DOCUMENT RESUME

ED 473 556 FL 027 564

TITLE	Revista de Investigacion Educativa, 2000 (Journal of Educational Research, 2000).
ISSN	ISSN-0212-4068
PUB DATE	2000-00-00
NOTE	701p.; Published twice per year. For Volume 17, see FL 026 653.
AVAILABLE FROM	Asociacion Interuniversitaria de Investigacion (AIDIPE), Dpto. de Metodos de Investigacion y Diagnostico en Educacion, Facultad de Pedagogia, Paseo del Valle de Hebron, 171, 2 planta, 08035 Barcelona, Spain. Web site: http://www.uv.es/aidipe/.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Revista de Investigacion Educativa; v18 n1-2 2000
LANGUAGE	Spanish
EDRS PRICE	EDRS Price MF04/PC29 Plus Postage.
DESCRIPTORS	*Educational Research; Elementary Secondary Education; Evaluation Methods; Foreign Countries; Higher Education; Learning Disabilities; Physical Education; Professional Development; Research Methodology; Self Concept; Special Education; Teacher Evaluation; *Teacher Student Relationship; Teaching Methods
IDENTIFIERS	*Spain

ABSTRACT

Articles in this volume focus on the following: teacher evaluation and quality management in education; steps toward a comprehensive and systematic staff evaluation; opinions of university students on teaching methods at science faculties; design of a scale to assess the ability to jump for the use in elementary school physical education; effects of age and gender differences in student-teacher relationships in an interactive cognitive-behavioral study designed to improve self-esteem; evaluation of a counseling program of prisoners for the reintegration into the workforce; statistical models to establish determining factors for choosing a university career and performance during the first year of teacher training; self-concept, self-esteem, and academic performance of students during the 4th year of Obligatory Secondary Education (Educacion Secundaria Educativo--E.S.O.); initial training of Secondary Education teachers. Current situation and outlook; methodology for teaching mathematics through problem solving: An evaluative study; and analysis and validation of a scale for measuring explorative behavior; evaluation of educational programs; evaluation of social programs; design and sub-designs in program evaluation; validity in program evaluation; data analysis in the evaluation of educational programs; optimizing the assessment of program evaluation; survey sampling and its problems in institutional program evaluations; evaluation and improvement of teaching at the university; evaluation of teaching by students within the National Plan of Evaluation of the Quality of the Universities; construction of an assessment instrument; evaluations of university training programs in a European and American framework; new trends in the evaluation of multicultural education programs; evaluation of computerized programs; evaluation of professional and continued education; evaluation of programs for the education of highly gifted students; evaluation of programs for high

Reproductions supplied by EDRS are the best that can be made
from the original document.

ability students; evaluation of special education programs for children with mental disorders; evaluation of programs focusing on children with disabilities. (AS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 473 556

Revista de Investigacion Educativa, 2000 (Journal of Educational Research, 2000)

Revista de Investigacion Educativa; v18 n1-2 2000

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

☐ Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. P. Sandin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

FL 027 564

BEST COPY AVAILABLE

Volumen 18, número 1, 2000

Revista de Investigación Educativa

Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE)

FL027564

ISSN: 0212-4068
Depósito Legal: MU-724-1996

Revista de Investigación Educativa

Volumen 18, número 1, 2000

Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE)

Director:
Eduardo Abalde Paz

Directora ejecutiva:
Mercedes Rodríguez Lajo

CONSEJO ASESOR:
Margarita Bartolomé
Nuria Borrell
Leonor Buendía
Iñaki Dendaluce
José Cajide
Narciso García
Fuensanta Hernández Pina
Jesús Jornet
Joan Mateo Andrés
Mario de Miguel
Arturo de la Orden
Carmen Jiménez Fernández
Antonio Rodríguez Diéguez
Francisco J. Tejedor

CONSEJO DE REDACCIÓN:
Manuel Álvarez
Rafael Bisquerra
Flor Cabrera
Inmaculada Dorio
Julia Victoria Espín
Pilar Figuera
Mª Ángeles Marín
Mercè Noguer
Mª Luisa Rodríguez
Mercedes Rodríguez
Mª Paz Sandín
Trinidad Donoso

SUSCRIPCIÓN E INTERCAMBIO CIENTÍFICO:
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Pedagogía
Paseo del Valle de Hebrón, 171, 2ª planta
08035 BARCELONA (Spain)

DISTRIBUCIÓN:
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Educación
Campus de Espinardo
Universidad de Murcia 30100
Tel. (968) 36 40 67

DISEÑO, MAQUETACIÓN E IMPRESIÓN:
Compobell, S.L. MURCIA

ISSN: 0212-4068

Depósito Legal: MU-724-1996

Revista de Investigación Educativa

Volumen 18, número 1, 2000

Editorial 3
Eduardo Abalde Paz

ESTUDIO MONOGRÁFICO

La evaluación del profesorado y la gestión de la calidad de la educación. Hacia un modelo comprensivo de evaluación sistemática de la docencia 7
Juan Mateo Andrés

TRABAJOS DE INVESTIGACIÓN

La opinión de los estudiantes universitarios sobre el método docente de las facultades de ciencias 37
Francisco José Sáez

Diseño de escalas para la evaluación de la habilidad motriz de salto para la educación física en Educación Primaria 47
Francisco Javier Castejón Oliva

Efectos de la diferencia de edad y sexo entre tutor y tutorado en una intervención cognitivo-conductual dirigida a mejorar la autoestima, basada en una estructura de tutoría recíproca 69
Miguel Moreno Moreno

Evaluación del programa de orientación para la inserción laboral en los centros penitenciarios 83
Gemma Filella Guiu

Modelización estocástica de la elección de carrera universitaria y de los resultados obtenidos en primer curso de las titulaciones de Pedagogía y Magisterio . 97
Gómez García J., Hernández Pina F., Buendía Moya F. y Carmona Martínez M.

Autoconcepto, autoestima y rendimiento académico en alumnos de 4º de E.S.O. Implicaciones psicopedagógicas en la orientación y tutoría 119
Miguel Ángel Broc Cavero

La formación inicial del profesor de Educación Secundaria. Situación actual y perspectivas de futuro 147
Antonio Pantoja Vallejo y Tomás J. Campoy Aranda

Metodología para la enseñanza de las matemáticas a través de la resolución de problemas: un estudio evaluativo 175
Teresa González Ramírez

Análisis y validación de una escala para medir la conducta exploratoria 201
Trinidad Donoso, Pilar Figuera y Mercedes Torrado

TRABAJOS METODOLÓGICOS

Criterios de validez en la investigación cualitativa: de la objetividad a la solidaridad 223
M. Paz Sandín Esteban

FICHAS RESUMEN DE INVESTIGACIÓN 243

E D I T O R I A L

En el IX Congreso Nacional de Modelos de Investigación Educativa: «Nuevas realidades educativas. Nuevas necesidades metodológicas» celebrado en la Universidad de Málaga se acordó, en la Junta de AIDIPE celebrada, que los compañeros del área MIDE de la Universidade da Coruña asumieran el compromiso de la celebración del próximo Congreso y, por primera vez, también la presidencia de esta asociación cayera sobre una persona de la misma Universidad. Desde aquí y, aprovechando este foro, quiero agradecer a todos los asistentes y especialmente a los socios de AIDIPE, su generosa confianza en un cargo de tan alta responsabilidad como esta, así como por la tarea encomendada.

Es, por el cumplimiento de los acuerdos de la AIDIPE en Asamblea General de AIDIPE, por la que me corresponde presentar este nuevo volumen de la Revista RIE con una enorme satisfacción. A lo largo de estos años, la Revista ha servido como medio de, por una lado, recogida de todos aquellos trabajos científicos presentados en las diferentes secciones de los Congresos celebrados y, por otro lado, como medio de comunicación de los diferentes profesionales de la Educación para dar a conocer aquellas investigaciones, experiencias y prácticas profesionales que se están desarrollando no sólo en nuestro país, sino que, con la vinculación con la European Educational Research Asociation (EERA), se están desarrollando en Europa. Es en esta línea, donde la Revista aúna esfuerzos, que en cierto modo creemos comunes, y que se plasman y vertebran en el eje común de la investigación educativa en todas sus secciones.

En el volumen 16, nº. 1 de 1998, mi precedesor en el cargo, el profesor Joan Mateo Andrés, reflexionaba sobre la necesidad de atender a las nuevas necesidades y de reconceptualización para integrar los nuevos conocimientos y metodologías haciendo eco de la necesidad de que nuestra asociación vertebre nuevas formas organizativas para arbitar nuevas vías de expresión, de pomover nuevos esfuerzos de pensamiento,... El Congreso al que anteriormente nos referíamos ha dado un paso de gigante en esta línea, pero, no cabe duda que queda mucho por hacer, pero no sin felicitarnos todos de la constante aportación que realizamos por la mejora de la calidad educativa.

Es, en aras de esta mejora por la que hemos de llegar al siguiente Congreso de AIDIPE, primero del milenio y que se celebrará en el 2001 en A Coruña, con el objetivo de tratar aquellos temas que más nos preocupan desde diferentes áreas científicas; en las que, Conferencias, Ponencias, Simposios, Comunicaciones y Posters sean motivo

de debate y enriquecimiento de nuestro conocimiento y de nuestro quehacer diario como profesionales de la educación.

Por ello, y desde aquí os invito a que con vuestras ideas, colaboración, trabajos y esfuerzos hagamos que el *X Congreso Nacional de AIDIPE*, que se celebrará en A Coruña del 19 al 21 de septiembre de 2001, cuyo título es «*Investigación y Evaluación en la Sociedad del Conocimiento*», tenga la repercusión social y académica de las anteriores ediciones.

Finalmente, en el número de RIE que se presenta podremos encontrar una serie de artículos que, con toda seguridad, serán de gran interés para el lector y que despertarán en él nuevas perspectivas teóricas y metodológicas.

Eduardo Abalde Paz
Presidente

ESTUDIO MONOGRÁFICO

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 7-34

LA EVALUACIÓN DEL PROFESORADO Y LA GESTIÓN DE LA CALIDAD DE LA EDUCACIÓN. HACIA UN MODELO COMPRENSIVO DE EVALUACIÓN SISTEMÁTICA DE LA DOCENCIA

Joan Mateo
Universidad de Barcelona

RESUMEN

Existe un renovado interés por el estudio de la evaluación del profesorado como ingrediente fundamental en la mejora de la escuela. Este interés proviene del convencimiento que detrás de cualquier mejora en la escuela tiene al profesor en su centro de gravedad.

En nuestro artículo tratamos de sintetizar la investigación y las conceptualizaciones existentes sobre la evaluación del profesorado, haciendo especial énfasis en su relación con la mejora de la escuela y la gestión de la calidad educativa.

Lo hemos estructurado en cuatro apartados. Uno introductorio y otros tres que tratan de explicar las condiciones y las características de un modelo comprensivo de evaluación sistemática de la docencia, así como indicaciones de cómo diseñarlo adecuadamente.

SUMMARY

There is renewed interest into the role of teacher evaluation as a fundamental aspect of school improvement. This interest comes from the realization that any significant improvement in schooling must have the teacher at his heart.

This article synthesizes current research and thinking about teacher evaluation related with school improvement and with educational quality management.

BEST COPY AVAILABLE

The article es organized into 4 parts: an introductory one and 3 that intend to explain the conditions and characterístics for a comprensive model of teacher evaluation system, and how to design it properly.

INTRODUCCIÓN

Existe, sin duda, un renovado interés por el papel que juega la evaluación del profesorado en la mejora de la escuela. De hecho, las comunidades educativas no pueden sustraerse a la creciente preocupación social en torno a como introducir en todos sus ámbitos los modelos de gestión de la calidad y es evidente que existe la convicción de que detrás de cualquier mejora significativa en la escuela subyace la actividad del profesorado. Por todo ello, establecer la conexión racional entre los vértices del triángulo: evaluación de la docencia-mejora y desarrollo profesional del profesorado-mejora de la calidad de la escuela, es fundamental para introducir acciones sustantivas de gestión de la calidad en cualquier institución educativa.

A través de este escrito trataremos de analizar y resumir el pensamiento y la investigación actual en torno al tema de la evaluación del profesorado y como se traduce en la práctica, tema al que hemos dedicado una importante parte de nuestro tiempo de investigación, de gestión y de práctica educativa Para facilitar la comprensión del mensaje he subdividido la aportación en cuatro apartados, que no son equivalentes en extensión pero si suponen cuatro momentos distintos en el complejo proceso que afrontamos:

- Una nueva cultura de la evaluación en el contexto de cambio del paradigma organizativo de la educación.
- Aspectos previos a la elaboración de un sistema de evaluación del profesorado.
- Fases en el diseño de un sistema de evaluación del profesorado.
- Evaluación y desarrollo profesional.

UNA NUEVA CULTURA DE LA EVALUACIÓN EN EL CONTEXTO DE CAMBIO DEL PARADIGMA ORGANIZATIVO DE LA EDUCACIÓN

1. Un marco social distinto para la educación

Como consecuencia de las profundas mutaciones que se están produciendo en las sociedades industrializadas, en sus economías y en la organización del trabajo, los sistemas educativos de los países europeos y de forma más generalizada los del mundo industrializado, se han visto obligados a introducirse en un verdadero cambio de paradigma organizativo.

La secular vinculación entre formación y mercado de trabajo que había conducido a la percepción de la educación como inversión ha quedado desdibujada ya que en la actualidad la acreditación educativa no garantiza un espacio en el mundo laboral y en una sociedad tan mercantilizada como la nuestra, donde la lógica del mercado se extiende sin reflexión ni crítica a todos los ámbitos, produce temor el pensar que

puede ocurrir si bajo el prisma de ser considerada la educación como un gasto social más, se le aplican las políticas restrictivas propias de los nuevos tiempos.

Como señala Tedesco, (2000, p. 82), «es necesario abrir los sistemas educativos a las prioridades de la sociedad actual y redefinir sus pactos con los agentes socializadores». Es evidente que el problema actual en los países occidentales no está tanto en la universalización de la enseñanza sino en garantizar a sus ciudadanos el acceso a una educación de calidad. Su principal prioridad no es crecer cuantitativamente sino gestionar correctamente la calidad de la educación y facilitar el acceso a todos los ciudadanos a una educación de mayor exigencia.

Sólo desde una enseñanza de calidad que dote al conjunto de ciudadanos de los instrumentos y las competencias cognitivas necesarias para un desempeño ciudadano activo, se puede tratar de soslayar los problemas de exclusión social asociados a la modernidad.

La educación debe atender los nuevos desafíos sobre los que la sociedad se está construyendo. Debe garantizar los cuatro pilares de Delors (1996): aprender a conocer, a hacer, a convivir y a ser. Debe, en un elemental ejercicio democrático, asegurar en definitiva a todos los ciudadanos el acceso a los circuitos en los cuales se produce y se distribuye el conocimiento socialmente más significativo.

2. Principios generales del nuevo paradigma

En cuanto a la orientación, la evolución de los países europeos en relación a los sistemas educativos, parece responder a los cuatro principios generales que exponemos a continuación (Hutmacher, 1999, p. 24), que dan soporte y explican la aparición de un nuevo paradigma organizativo:

a) *La primacía de las finalidades*, que exige que las decisiones y la acción se orienten de una manera prioritaria conforme a la voluntad de alcanzar los objetivos establecidos, en todos los niveles del sistema, desde el aula hasta los niveles politico-estratégicos.
b) *El principio de imputabilidad (accountability)*, según el cual los agentes de todos los niveles del sistema valoran y deben responder sistemáticamente del grado de logro de los objetivos.
c) *El principio de subsidiaridad*, según el cual las decisiones se han de tomar en el mismo nivel donde se produce su implementación, y únicamente se delegan a un nivel superior de responsabilidad si los objetivos no es posible obtenerlos de otra manera.
d) *El principio de auto-organización y de desarrollo continuo*, según el cual los sistemas y sus componentes nunca están absolutamente fijados, sino que están en continua transformación, bien sea por la propia retroalimentación o bien reforzados por el entorno.

3. Hacia una nueva cultura de la evaluación

Este cambio implica una extensión importante de los mecanismos de evaluación a todos los niveles del sistema y entre ellos ocupa un lugar preeminente la escuela. Por otro lado esta extensión se acompaña de un cambio de lugar de la cultura de la evaluación que ya no se orienta tanto hacia la sanción, la clasificación y la selección como hacia aportar una información razonada y razonable dirigida a la orientación de la gestión de la mejora educativa.

Esta información actúa como retroalimentadora de la acción y apunta a estimular la reflexión y la búsqueda de las soluciones más eficaces y eficientes por tal de responder a las expectativas exigentes y a menudo contradictorias dirigidas a los sistemas educativos desde la sociedad.

En el marco de este cambio los sistemas educativos se enfocan de otra manera y aunque continúe su expansión cuantitativa, se da más importancia a la calidad: calidad al servicio de los individuos, de los alumnos y calidad desde el punto de vista de criterios sociales, políticos, culturales y económicos más globales.

Pero, la exigencia de calidad necesita ser concretada y discutida, lo que significa una reorientación de las finalidades de los sistemas educativos, en los que se estrecha la asociación entre conocimientos y competencias intelectuales y significa también analizar el papel que juegan cada uno de los agentes educativos en función de la globalidad de la misión con la intención de incidir en la mejora de la gestión de la calidad.

Las políticas educativas de la mayor parte de los países europeos tienden a pasar de una lógica centrada en los «inputs» a una lógica centrada en los logros y en los procesos que los producen y que tratan de explicarlos convincentemente.

En el nuevo paradigma organizativo emergente, la evaluación se convierte en parte de los procesos de control y de desarrollo de la propia calidad y sirve también para la rendición de cuentas. Reviste una gran importancia estratégica y está al servicio de la progresión hacia la realización de las finalidades del sistema y del conjunto de los componentes que están al servicio del desarrollo hacia la calidad.

Lo centros representan una pieza clave en la nueva arquitectura organizativa, se convierten en ámbitos de reflexión y de innovación pedagógicas, su estatus en la gestión de la calidad se revaloriza gracias a que se constituyen en comunidades educativas autónomas y responsables, encargadas de realizar a nivel local el proyecto educativo global fijado para el conjunto del sistema, pero con una capacidad de autonomía y una responsabilidad amplias por lo que hace a las modalidades de actuación que le permiten profundizar y generar un producto educativo propio que resalte su propia identidad.

Es en este marco donde debe insertarse la evaluación de la docencia y es a partir de ella que debe analizarse y explicarse la del profesorado. No es posible dotar de significación a la evaluación del profesorado sino es insertada en los procesos de mejora de la calidad de la docencia en el contexto del centro educativo.

ASPECTOS PREVIOS A LA ELABORACIÓN DE UN SISTEMA DE EVALUACIÓN DEL PROFESORADO

Procesualizar y establecer las bases y la estructura de un sistema evaluativo para el profesorado no es una tarea sencilla, prueba de ello está en las dificultades que se dan a nivel de la mayoría de los países occidentales para homologar modelos y configurar praxis completas de evaluación. En general sólo se ejecutan intervenciones parciales y con poca continuidad. Las reflexiones que exponemos a continuación son el producto de la investigación y de la reflexión conceptual y tratan de establecer los elementos previos clave para el establecimiento de un sistema evaluativo del profesorado.

1. Las expectativas del centro versus las necesidades del profesorado

Tal como hemos señalado, el problema de fondo no es la evaluación, sino como a partir de ella conseguimos gestionar más eficientemente la calidad. Para ello deberemos establecer en primer lugar los elementos básicos que sustentan un modelo de evaluación del profesorado y como estos se relacionan con los procesos de mejora de los centros.

Es evidente que un ingrediente fundamental en los logros de calidad de un centro, lo constituye la calidad de la docencia producida por su profesorado, pero... ¿hasta qué punto son coincidentes los intereses de uno y otro? El paso previo para establecer un modelo de evaluación del profesorado es resolver el sistema relacional entre ambos conjuntos de necesidades.

Se admite que en una organización sana, la relación entre el centro y el profesorado ha de ser dinámica y el principio de que lo que es bueno para el centro es bueno para el profesorado y viceversa debe prevalecer. Este tipo de relación sinérgica favorece la posibilidad de que ambos alcancen sus propósitos, equilibrar las necesidades individuales con las expectativas institucionales es esencial para generar ambientes de trabajo productivos (March y Simon, 1993). Este ha de ser a nuestro juicio el primer objetivo a cubrir en cualquier estrategia evaluativa.

No es posible plantearse un verdadero proceso de mejora de la calidad que ignore las demandas individuales. Tal como señala Fullan (1991, p. 349) «Combinar el desarrollo individual e institucional genera sus tensiones, pero el mensaje es evidentemente claro. Nadie puede obtener uno sin el otro».

Todo modelo de evaluación del profesorado deberá plantearse como primer paso el contar con la participación activa del mismo en su diseño. Sólo desde la participación es posible contemplar exhaustivamente las necesidades del profesor y tratar de compatibilizarlas con las del centro, de lo contrario, el proceso culturalizador que toda acción evaluativa debe contener estará ausente, la evaluación se convertirá en un proceso generador de incultura evaluativa.

Los modelos participativos se caracterizan por posibilitar e incentivar la intervención y la implicación de las personas y los grupos que conforman la organización en la toma de decisiones que afectan a la tarea evaluativa a realizar: delimitación de objetivos, establecimiento de planes de acción, distribución de funciones, análisis de la

información, establecimientos de juicios de valor, generación de decisiones de mejora, seguimiento, etc.

La participación del profesorado en los procesos de evaluación no se ha de entender únicamente como una estrategia de gestión, sino que presenta un valor formativo intrínseco.

Desde la implicación resulta más viable el aprovechar algunos elementos educativos contenidos en la evaluación, como es su capacidad de desarrollar los procesos de autorregulación personales o de grupo.

La participación facilita la coordinación de las actuaciones, posibilita la construcción de una visión unitaria del centro y la creación de una cultura institucional compartida.

Sólo activando la participación seremos capaces de crear un clima en el que sea factible el conjugar las necesidades a que aludíamos al principio de este apartado. Desaprovechar esta función supone comprometer gravemente, tal como señalaba Fullan, la gestión correcta de la calidad.

2. Propósitos básicos en un proceso de evaluación del profesorado

Clarificadas las expectativas del centro, paso previo para establecer su misión y de ahí derivar su definición de calidad y compatibilizadas con las necesidades del profesorado procede pasar a establecer el propósito evaluador.

Los dos propósitos más frecuentemente citados por la mayoría de autores para la evaluación del profesorado son la petición de responsabilidades y la mejora de la actividad docente (McLaughlin y Pfeifer, 1988, Duke 1990, Mateo et al. 1996, 1999). Sin embargo no nos parecen suficientes, así expresados, desde la perspectiva de la gestión de la calidad.

La calidad educativa de un centro no depende directamente de la calidad de la actividad docente del profesor ni de que se le exijan responsabilidades. Es la calidad de la docencia el núcleo básico impulsor de la calidad del centro y la calidad personal del profesor y la de su actividad no tienen que estar necesariamente relacionadas con ella.

Para que la calidad del profesor este relacionada con la de la docencia será preciso establecer previamente el marco conceptual en el que se fundan dinámicamente los intereses y necesidades individuales e institucionales. No es posible determinar si una acción del profesor es realmente de calidad sin acudir al marco de referencia que ofrece la visión global de lo que se entiende en un centro determinado por calidad de la docencia.

Hemos de desarrollar una concepción de la evaluación del profesorado en la que se desarrolle simultáneamente la mejora del profesor y la de la institución. Este modelo comprensivo de la evaluación del profesorado estaría enraizado en dos amplios propósitos (Stronge, Helm y Tucker, 1995):

a) *Orientado a los resultados,* contribuyendo a la consecución de los objetivos personales del profesor y a los de la misión del programa, del centro y de la organización educativa en su totalidad. Debería también proveer de información clara

y fiable para la petición de responsabilidades por la actuación docente en función de esos objetivos.

b) *Orientado a la mejora,* contribuyendo al desarrollo personal y profesional del profesor así como a la mejora del centro.

Todo centro debe tener explicitado su modelo de calidad que se desprende de su misión como organización. Debe tener perfectamente establecidos los objetivos a lograr y cuales corresponden a la docencia. La calidad de la actividad docente del profesor se establece a partir del modelo de calidad para la docencia en el centro. Sólo desde la calidad de la docencia pueden entenderse los objetivos que debe alcanzar el profesorado en su actividad docente y las orientaciones metodológicas que debe aplicar y consecuentemente establecer el correspondiente modelo de evaluación y sólo desde ella pueden orientarse procesos de mejora del profesorado que tengan significado para el centro y para la persona simultáneamente.

3. Acciones básicas para promover el proceso de evaluación del profesorado

Llegados a este punto y si queremos alcanzar una evaluación del profesorado de alta calidad, el sistema evaluativo deberá construirse a partir de un equilibrio dinámico entre la mejora de la escuela y la mejora del profesorado y para ello será necesario tener en cuenta las siguientes acciones (Stronge, 1997):

a) Establecimiento de objetivos mutuamente beneficiosos.
b) Enfatización de la comunicación sistemática.
c) Creación de un clima favorecedor de la evaluación.
d) Garantía en la aplicación técnica del sistema evaluativo.
e) Uso de múltiples fuentes de datos.

a) Establecimiento de objetivos recíprocamente beneficiosos

Los objetivos que van a constituir las metas a conseguir deben ser valorados y percibidos como importantes tanto por los profesores como por la institución. Olsen y Bennet (1980) describen la actividad evaluativa como un proceso asociativo entre la organización y sus empleados, un proceso social que produce como resultado una aproximación participativa a la evaluación. Es en el marco de esta relación mutuamente beneficiosa que se produce en el ámbito laboral desde donde debería guiarse el desarrollo e implementación de un sistema evaluativo relevante y mutuamente beneficioso.

b) Establecimiento de un proceso sistemático de comunicación

Los sistemas de evaluación del profesorado deben construirse a partir de un proceso compartido de comunicación donde todos los aspectos clave deben ser tratados informativamente de forma correcta. Es muy importante establecer los momentos informativos tanto de aquellos aspectos que son públicos por naturaleza como aque-

llos que requerirán de un tratamiento informativo más privado. La transparencia de los procesos y la posibilidad de interacción se fundamentan en políticas eficientes de información, por tanto cada aspecto clave del proceso evaluativo deberá ir acompañado del correspondiente acto informativo y este debe llevar asociadas las condiciones de garantía de que se realiza en condiciones óptimas para favorecer un verdadero proceso comunicativo.

c) Cuidado en la aplicación técnica del sistema de evaluación

Un sistema evaluativo técnicamente correcto no garantiza una evaluación efectiva, pero evidentemente un sistema incorrecto si que garantiza que nunca llegue a serlo. Posteriormente entraremos con más detalle en la estructura y características que debe contener un sistema evaluativo y en la instrumentación disponible pero sí que queremos hacer aquí un especial énfasis en la necesidad de que cada uno de estos aspectos que conforman la aplicación técnica de los procesos evaluativos se realice de acuerdo a las condiciones técnicas de calidad y precisión exigidas por las más elementales normas de la evaluación educativa.

d) Uso de fuentes de datos múltiples

Aunque el método más usado en la evaluación del profesorado es el que podríamos denominar como clínico, basado en una secuencia que se inicia con una preconferencia, sigue con observación en el aula y acaba con la postconferencia (E.T.S. 1988), los sistemas modernos de evaluación del profesorado se caracterizan por documentar la actividad docente a partir de múltiples fuentes de información (Glatthorn, 1984; Conley, 1987; Harris, 1987 y Mc Greal 1988).

El uso de fuentes múltiples de datos para la evaluación del profesorado ofrece numerosas ventajas respecto de la evaluación basada en una sola fuente, citamos entre ellas (Stronge, 1997):

- Producción de un marco de la actuación evaluada mucho más rica y contextualizada.
- Recogida de datos en situaciones más naturales.
- Integración en la evaluación de fuentes primarias y secundarias de datos.
- Seguridad de una mayor fiabilidad en la documentación de la actuación docente.
- Documentación respecto de la actuación docente más cercana a la realidad.
- Unas bases más legalmente defendibles respecto de las decisiones evaluativas.

La integración de múltiples fuentes de datos en los procesos de evaluación del profesorado aporta una imagen mucho más real de la actividad desarrollada por el profesor y provee de una plataforma mucho más poderosa sobre la que construir planes de mejora realistas que la que se nos ofrece a partir de una sola fuente, como pudiera ser la observación en el aula. Usar apropiadamente múltiples fuentes en un proceso de evaluación potencia enormemente todas sus posibilidades.

e) Creación de un clima adecuado para la evaluación

La evaluación sólo puede resultar absolutamente eficiente si se lleva a cabo desde un clima adecuado donde la confianza entre las partes implicadas y la honestidad y la transparencia en las actuaciones sean la norma y no la excepción.

La consecución de estos climas se posibilita en la medida en que se sigue los pasos establecidos: participación en el establecimiento de los objetivos, comunicación continua y fluida en las diferentes fases del proceso, precisión en la aplicación de las técnicas, honestidad en los juicios y enfoques proactivos en los procesos de mejora.

No se puede ocultar que en ocasiones los procesos evaluativos no se llevan a término en las mejores condiciones en cuanto a clima. En estos casos la prudencia debe presidir las actuaciones y en muchos casos deberemos abandonar algunas de ellas por tal de no quemar las posibilidades de mejora. Es preferible reducir los ámbitos de actuación y tratar de avanzar lentamente procurando que sea el propio proceso el que nos vaya abriendo el camino para acometer tareas de mayor dificultad.

FASES EN EL DISEÑO DE UN SISTEMA DE EVALUACIÓN DEL PROFESORADO

Distingo cuatro grandes fases en el proceso de diseño de un sistema de evaluación del profesorado:

1. La fase de ideación.
2. La fase de desarrollo.
3. La fase de implementación.
4. La fase metaevaluativa

1. Fase de ideación

Durante esta fase se deben resolver básicamente los siguientes aspectos:

a) Identificación del sistema de necesidades.
b) Derivación de los objetivos institucionales.
c) Definición del modelo de calidad para la docencia.

a) Identificación del sistema de necesidades

Cada centro tiene sus necesidades específicas que lógicamente han de estar relacionadas con el contexto social y cultural y que deben ser alcanzadas por la acción colectiva de todo el equipo educativo.

Un examen minucioso de las necesidades de la escuela constituye el paso previo para clarificar la misión y el propósito educativo del centro, ayuda a definir el proyecto educativo y a clarificar los objetivos institucionales.

b) Derivación de los objetivos institucionales

Determinar las necesidades de la organización es un prerrequisito absolutamente necesario para establecer si los restantes pasos del proceso evaluativo son relevantes en términos de la misión del centro y de las demandas de petición de responsabilidades provenientes de la sociedad (Connellan, 1978, Castetter, 1981, Patton, 1986 y Goodale, 1992).

De él se derivan los objetivos institucionales y su conocimiento favorece su operativización y juntamente con el análisis de la capacidad del sistema vehicula el establecimiento del grado de consecución que debemos pretender. Todo ello constituirá las bases del modelo de calidad de la docencia del centro. Explicitar el modelo de calidad de la docencia posibilita la inserción de la evaluación del profesorado en los procesos de gestión de la calidad de la institución.

c) Definición del modelo de calidad para la docencia

Tal como hemos señalado anteriormente la operativización de los objetivos institucionales y la clarificación de su nexo relacional con las necesidades del centro, permiten juntamente con el análisis de los recursos disponibles el establecer un modelo de calidad de la docencia significativo y sostenible para la institución.

El modelo de calidad de la docencia orienta y sustantiva el proceso de evaluación del profesorado. Es el vínculo entre la evaluación del profesor y la mejora de la calidad del centro. Su definición debe facilitar la descripción y desarrollo de las funciones del profesor y de sus responsabilidades y orientar el establecimiento de los «estándares» de la actuación profesional del mismo.

2. Fase de desarrollo

Bajo esta fase contemplamos dos acciones:

a) Determinación de las funciones y responsabilidades del profesorado.
b) Desarrollo de los criterios de evaluación y de sus indicadores.
c) Fijación de los «estándares» para la valoración de la actividad docente.

a) Determinación de las funciones y responsabilidades del profesorado

Hemos de partir de la premisa que sólo a partir de modelos de docencia bien establecidos y definidos es posible establecer descripciones apropiadas y ajustadas de las funciones y responsabilidades del profesorado (Olsen y Bennett, 1980 y Scriven, 1991).

Todo sistema de evaluación del profesorado exige establecer con precisión las áreas de responsabilidad del profesor en el contexto institucional y sus funciones más importantes. Estas áreas han de servir de marco para establecer con la mayor precisión posible las obligaciones y actividades a cumplir por el profesor y cual es su sentido en el contexto del desarrollo general de la docencia en el centro.

Cabe añadir que los sistemas evaluativos más desarrollados incorporan sistemas de indicadores en relación con la acción docente del profesor. Tratan de capturar mediante el uso de muestras de indicadores aspectos esenciales de la profesión docente. Parten del principio de que sólo a partir del indicador es posible objetivar la acción y la reflexión evaluativa.

b) Desarrollo de los criterios de evaluación y de sus indicadores

Una vez establecidas las funciones y áreas de responsabilidad básicas del profesorado y el modelo de calidad de la docencia, podemos derivar factores asociados a las mismas que nos servirán de criterios para evaluar. Una variable criterio nos provee de una dimensión general dentro de la que será posible valorar o juzgar el nivel de realización alcanzado por el sujeto (Wheeler y Haertel, 1993). El área de responsabilidad nos orienta acerca de las variables criterio más significativas y el modelo de calidad direcciona el sentido de la evaluación.

Los indicadores son observables que nos permiten cuantificar las variables criterio y facilitan los procesos de objetivación de la evaluación.

Criterios e indicadores constituyen piezas muy importantes en los procesos de concreción de los sistemas evaluativos más desarrollados.

c) Fijación de «estandars» para la valoración de la actividad docente del profesor

Fijar «estándares» implica determinar los niveles de actuación del profesor considerados como aceptables. Evidentemente para poder proceder a establecer los referentes que nos permitirán determinar si una determinada actuación docente ha sido realizada en términos de calidad aceptable de acuerdo al modelo de calidad establecido para el centro, es preciso haber recorrido al menos mínimamente los elementos hasta ahora establecidos.

Únicamente si tenemos bien establecidos los objetivos institucionales, su naturaleza evaluativa y el marco general de la docencia, será posible pensar en el establecimiento de criterios y niveles de valor previamente determinados. Contar con un sistema de indicadores bien establecido facilita enormemente la tarea, ya que la objetivación de la información permite el uso de los referentes normativos y criteriales en el proceso de estandarización de las valoraciones. La estandarización puede llevarse a término a cualquier nivel de análisis, aunque obviamente cobra significados distintos según sea el mismo. La unidad referencial esencial en la evaluación del profesorado es, a nuestro juicio, el centro educativo.

Cabe añadir, por otro lado, que la determinación de los «estándares» para cualquier actividad evaluativa, incrementa la posibilidad de transparencia y limpieza de los juicios valorativos y la defensa legal del sujeto frente a posibles arbitrariedades. (Phi Delta Kappa National Study Committee on Evaluation, 1971, Committee on Standars for Educational Evaluation, 1988).

3. Fase de implementación

Para el desarrollo de esta fase consideramos tres momentos:

a. Documentación de la actividad profesional del profesor.
b. Emisión de los juicios valorativos respecto de la docencia.
c. Uso intensivo y extensivo de la información: la derivación de las propuestas de mejora.

a) Documentar la actividad profesional del profesor

Por documentar, entendemos el proceso de recogida de suficientes evidencias acerca de la actividad profesional del profesor que de soporte a la posterior evaluación y justifique racionalmente las tomas de decisiones personales e institucionales basadas en la evaluación.

Tal como ya hemos señalado anteriormente, en la evaluación moderna los procesos de documentación deben relacionarse con múltiples fuentes y estrategias de recogida de datos, en ningún caso deberemos construir un proceso evaluativo basado en una única fuente de información (Conley, 1987, Ellet, 1987, Harris, 1987, Mc. Greal, 1988, Helm, 1994).

Entre los sistemas más utilizados en la actualidad, para evaluar el profesorado, tenemos (Stronge, 1997): la observación en el aula, la opinión de los padres y alumnos, juicios valorativos emitidos por pares, el rendimiento en los aprendizajes de los alumnos, el autoinforme y el portafolios.

La observación en el aula

En esta metodología se parte de la premisa de que observar al profesor en acción es la mejor forma de reunir información acerca de su efectividad docente. La visita al aula aporta evidencias muy difíciles de conseguir de otra forma como son: el clima en el aula, la naturaleza y calidad de las interacciones alumnos-profesor, permite observar los procesos de instrucción directamente y el funcionamiento general de la clase (Everston y Holley, 1981).

Sin embargo tiene también limitaciones, así Scriven (1981) nos señala la baja fiabilidad del procedimiento fundamentándose en las siguientes razones:

- El limitado número de observaciones y la brevedad de su duración no permiten la generalización.
- Hay un evidente peligro de que el evaluador focalice su atención guiado por sus intereses, reflejando, consecuentemente, más sus propios puntos de vista que la realidad del aula.
- La pobreza de los sistemas de registro propician que la mayoría de evaluadores se dejen influir por sus preconceptos.
- Cualquier interrelación entre observador y observado introduce factores de confusión.

- La propia visita altera el comportamiento del profesor y de los estudiantes de forma que reduce la posibilidad de que el evaluador realmente observe una muestra representativa de la docencia.

Otros elementos adicionales que limitan la fiabilidad de la acción observadora, son las que señalan Stronge, Helm y Tucker, 1995:

- La naturaleza artificial que entraña toda observación programada.
- Limitaciones por causa de la focalización en cualquier observación.
- La infrecuencia de las observaciones. (Si no se quiere alterar el ritmo de la clase, ¿cuántas visitas se pueden programar?)
- El hecho de que sólo una porción del repertorio total de las responsabilidades propias del profesor pueden ser observadas en una sola, o en unas pocas observaciones.
- La realidad de que existen muchas responsabilidades del profesor que se ejecutan fuera del aula (El trabajo en equipo, desarrollo colectivo de materiales, juntas de evaluación, etc.).

Independientemente de que la observación del trabajo del profesor en el aula pueda representar una aportación significativa y vital para un sistema comprensivo de evaluación del profesorado, es evidente que es insuficiente por sí sola. La mayoría de trabajos de investigación realizados en este sentido han mostrado sus debilidades (Root y Overly, 1990; Wood, 1992; Frase y Streshly, 1994) y en la línea de lo mencionado anteriormente los investigadores apuestan por el uso combinado de las diferentes estrategias (Stronge, 1997).

La opinión de los padres

En las últimas dos décadas se han llevado a término numerosos estudios analizando los beneficios producidos a partir de la inclusión de la opinión de padres en los procesos de evaluación del profesorado (Mager 1980; Becher 1984; Epstein, 1985; Mark y Shotland 1985). Parece obvio que una buena relación entre padres y profesores y una comunicación efectiva casa-escuela, ha de repercutir en la calidad de la educación.

Potenciar la participación de los padres en los procesos de evaluación del profesorado, permite analizar perspectivas y conocer puntos de vista y aspectos imposibles de obtener mediante otras fuentes. Epstein 1984, en la p. 4 señala:

> «(Los padres) aportan información acerca de cómo el profesor interactúa con el alumno y la familia, responde a las necesidades de los estudiantes, asigna retos apropiados en libros y trabajo de casa (todos ellos considerados como indicadores de efectividad en la docencia). Los padres están legítimamente habilitados para, juntamente con otros jueces, valorar adecuadamente la efectividad de los profesores y los programas escolares».

En el proyecto Faucette (1995), los propios profesores expresaron su sorpresa al descubrir la cantidad y calidad de información que los padres poseían sobre sus prácticas docentes en la clase. En la investigación mencionada aceptaron que los datos

que poseían eran valiosos, precisos y útiles y que podían ser integrados en sus procesos evaluativos y utilizados eficazmente en su desarrollo profesional como docentes y en la mejora de la docencia.

No queremos ocultar, sin embargo, las dificultades que entraña la utilización de esta fuente de información. Los padres con todo, no son profesionales de la educación, y sus juicios pueden reflejar sus prejuicios, su visión condicionada por sus propios hijos y también su desconocimiento del área de conocimiento. El profesor por otro lado puede sentirse acosado y como colectivo profesional tampoco es sencillo que acepten ser evaluados por personas ajenas al mismo (preguntémonos que opinarían el colectivo de arquitectos, por poner un caso, si en su evaluación profesional participaran sujetos ajenos a la profesión y no los pertenecientes a su colegio profesional).

Aún así, entendemos que la evaluación tiene como finalidad fundamental el aportar información, desarrollar capacidades y potenciar actitudes que puedan insertarse en los procesos de gestión de la calidad de la educación (así lo estamos defendiendo desde el principio de nuestro escrito), y las percepciones de los padres pueden en ese sentido aportar datos enriquecedores y colaborar en el desarrollo de actitudes de interrelación y mutuo conocimiento absolutamente necesarias para producir una educación de calidad.

No será de recibo extrapolar los datos hacia el análisis de aspectos que no correspondan y en ningún caso la participación evaluativa de los padres podrá entrar en el terreno de la petición de responsabilidades.

Su aportación debe servir fundamentalmente para aportar percepciones distintas, enriquecedoras de la visión comprensiva a la que aspiramos y sobre todo para moldear nuevas y muy positivas actitudes.

Como cierre de este apartado, bástenos citar del estudio de Faucette (1995), y a modo de ilustración de lo que tratamos de expresar, la reacción recogida de uno de los padres cuando después de participar en el proceso evaluativo expresa lo siguiente: «Estoy muy orgulloso que mi hijo vaya a una escuela donde los educadores valoren mi opinión. Doy la bienvenida a esta maravillosa oportunidad que me han brindado de ayudar al profesor de mi hijo».

La opinión de los alumnos

Incorporar la opinión de los alumnos en los procesos de evaluación del profesorado constituye uno de los tópicos más analizados por los estudiosos del tema. Uno de los argumentos más utilizados para justificar su inclusión se centra en que ellos son los consumidores primarios de los servicios del profesor.

Son, sin duda, los más directos receptores del proceso de enseñanza-aprendizaje y se encuentran en una posición priviligiada para proporcionar información sustantiva acerca de la efectividad de la docencia, más aún, los alumnos son los únicos que tienen información directa del tipo, naturaleza y calidad de las prácticas docentes que se realizan en el aula, así Follman (1992), señala, que no hay otros individuos o colectivos que tengan una experiencia directa con el profesor, en lo que se refiere a la docencia, que les supere en extensión, amplitud y profundidad de exposición.

Sin embargo la mayoría de los investigadores admiten restricciones en la recogida, en la interpretación y en el uso de la información proporcionada por los alumnos. En primer lugar parece claro que la práctica evaluativa involucrando alumnos debe restringirse a los de mayor edad, de hecho en nuestra realidad educativa y en la de la mayoría de países occidentales, sólo en los ámbitos universitarios constituye su uso una práctica extendida, aunque algunos estudios (Haak, Kleiber y Peck (1972), sugieren que pueden utilizarse a partir de la secundaria.

A nuestro juicio y a pesar del valor que todo el mundo adjudica a esta fuente de información, consideramos que la reflexión evaluativa de los alumnos respecto de sus profesores debe inscribirse en el marco exclusivo de la evaluación formativa del profesor y en su enfoque debe primarse el acentuar las posibilidades que encierra como instrumento para mejorar la relación alumno-profesor.

Que el profesor conozca las percepciones del alumno sobre su práctica docente y su actuación personal en el aula, independientemente de su valor formativo, puede ser una plataforma de interacción y acercamiento entre los alumnos y el profesor de primerísima magnitud. Los instrumentos utilizados en la recogida de información deberán estar construidos de manera que faciliten el acceso a aquel tipo de información que más ayude a ese propósito y deberá dejarse para otras estrategias aquella que pueda generar situaciones de conflicto, en la que la resolución no es sencilla, dada la natural asimetría existente entre el profesorado y el alumnado.

Finalmente cabe añadir que la información obtenida a partir de la opinión del alumnado, es también de un incalculable valor cuando se cruza con otras obtenidas mediante otros medios (p. ej. la de los padres).

Evaluación a partir del juicio emitidos por los pares

Al margen de la importancia y el valor que asignemos a la información obtenida a partir de las opiniones de los padres y de los alumnos, es un hecho comunmente aceptado que la información así obtenida ofrece una visión parcial de la docencia. Si verdaderamente queremos obtener una imagen precisa y completa de la misma, con la intención de gestionar su mejora y orientar las decisiones tanto personales como institucionales de una manera inteligente y ajustada, no tenemos más remedio que introducir en el sistema evaluativo información proveniente de los pares.

En los procesos modernos de mejora de la calidad y en el marco de la empresa, se ha adoptado últimamente en el mundo anglosajón el concepto denominado «valoración de 360 grados» (Stronge y Ostrander 1997, p. 150), consistente en retroalimentar positivamente la actividad de un empleado a partir de recoger información respecto de su actividad proveniente de todas las personas que le rodean y que de una manera u otra interactúan con él.

Los expertos consideran este concepto como paralelo al movimiento que persigue la mejora continua de la actividad docente a partir de usar la retroalimentación proveniente de los pares, como aportación fundamental en los sistemas de evaluación de la docencia.

Dada la naturaleza colaborativa de la enseñanza y la educación moderna, sería absurdo imaginar un proceso de evaluación que ignorara a los pares como fuente vital para la retroalimentación de la mejora de la calidad de la actividad docente. La evaluación por pares expresada de forma colegiada constituye sin duda, una fórmula extraordinaria de desarrollo profesional y de refuerzo positivo para los enseñantes.

Deberemos sin embargo introducir algunas precisiones básicas si queremos que realmente este principio sea aplicado adecuadamente. En principio en el contexto de un sistema de evaluación no hay que confundir el concepto de retroalimentación por los pares con el de evaluación por los pares.

La información aportada por los pares formará parte como la proveniente de una fuente más en el marco general del sistema de evaluación comprensivo de la docencia. En ningún caso se derivarán juicios valorativos directos desde esta información.

Por todo ello, la información que se solicita de los pares nunca estará basada en juicios valorativos personales del compañero o en información subjetiva basada en la impresión (o en el desconocimiento) respecto de su actuación, sino en información objetivamente valorada respecto de aquellos elementos de su actividad, que los compañeros conocen directamente y que debrán expresar de tal manera que permite incidir en su mejora. Sólo así la información de los pares introduce elementos sustantivos para la retroalimentación positiva del profesor y le aporta verdadero valor añadido.

Si los pares participan en la aportación de información para la evaluación sistemática de la actividad docente de los compañeros, habrá que tener muy presentes los siguientes extremos:

- *Usos de información basada en hechos*. La más importante consideración en la recogida de datos es que esta esté focalizada fundamentalmente sobre descripciones factuales respecto de la actuación docente de la persona a evaluar. Para ello es importante invitar como proveedores de datos únicamente a aquellos pares que tengan conocimiento directo de la actividad del profesor a evaluar. (Stronge y Helm, 1991).
- *Consideraciones en torno a la validez y a la fiabilidad*. Una forma de incrementar la validez del estudio es preguntar única y exclusivamente cuestiones a la persona interrogada, que por razón de su conocimiento directo esté en condiciones de contestar, evitando otro tipo de consideraciones. En cuanto a la fiabilidad, cabe recordar que cuanto mayor sea la muestra de personas interrogadas mayor será la misma, por tanto convendrá incluir suficientes sujetos en los procesos evaluativos por pares.
- *Confidencialidad*. La confidencialidad debe garantizarse para todo el proceso evaluativo, de lo contrario difícilmente conseguiremos que la información aportada sea emitida sin restricciones de algún tipo.
- *Procedimientos de muestreo*. Un estudio en el que la muestra elegida para llevarlo a cabo no garantice una importante representatividad, constituye un estudio absolutamente inútil.

El rendimiento en el aprendizaje de los alumnos

No debemos olvidar que el propósito directo de la enseñanza es el aprendizaje y que la escolarización debería asegurar que cada nueva generación de alumnos acumule los conocimientos y destrezas necesarios para desenvolverse solventemente al llegar a la edad adulta ante las demandas que marca la sociedad. A tenor de esta afirmación, incluir el rendimiento de los alumnos como una fuente más en los procesos de evaluación del profesorado parecería del todo razonable.

Sin embargo la idea de ligar estrechamente el aprendizaje de los alumnos con la tarea del profesor o de la escuela no es, a pesar de los numerosos trabajos en esa dirección (Scriven, 1988; Millman y Sykes, 1992; Smylie,1994; Webster, 1994), una idea totalmente aceptada. Los problemas técnicos y políticos planteados al intentar relacionarlos significativamente han impedido, hasta ahora, el orientar de forma generalizada los procesos de evaluación del profesorado sobre el rendimiento en los aprendizajes de los alumnos (Berk 1988).

Aún así y en este marco comprensivo que venimos desarrollando, parece que las nuevas líneas de trabajo vuelven sobre este tópico, sólo que desde puntos de vista distintos a los mantenidos tradicionalmente por los defensores de incluir el rendimiento de los alumnos como objeto de evaluación de la tarea docente.

Se trata no tanto de valorar la tarea del profesor sobre el rendimiento absoluto de los alumnos, sino tratar de hacerlo sobre el valor añadido aportado por la docencia. Ello implica por un lado establecer el punto de partida del rendimiento anterior, valorando no tan sólo las ganancias de rendimiento sino analizándolas a la luz que aporta también el potencial de aprendizaje observado en los alumnos a lo largo del tiempo (obtenido mediante el análisis de los registros de su evolución en los últimos cursos), así como también de cruzar esta información con otras de tipo contextual que nos permita racionalizar y desagregar que parte de las ganancias obtenidas responden a una actividad docente de calidad y cuales son atribuibles a otras causas.

Por otra parte y tal como ya hemos expresado anteriormente, la información así obtenida formará parte de ese registro acumulativo de evidencias múltiples que nos ayuda a dibujar con la mayor precisión posible y con el mayor número de matices la realidad de la docencia ejercida por el profesorado. Al tiempo debemos nuevamente recordar que incluir en la reflexión del profesorado información de este estilo respecto del rendimiento es fundamental en los proceso de mejora personal y colectiva y aporta datos decisivos a la institución para la gestión de la calidad de la docencia.

El autoinforme

El autoinforme constituye un proceso por el que el profesor efectúa juicios acerca de la adecuación y de la efectividad de su propia actividad, con el objeto de establecer las bases que faciliten la mejora.

En estos procesos de autorreflexión es el propio profesor el que debe seleccionar, recoger, interpretar y juzgar la información referida a su propia práctica. Aquí es el profesor el que establece los criterios y fija los estandars para juzgar la adecuación de

sus creencias, conocimientos, habilidades y efectividad respecto de su actividad docente. Finalmente es también él quien decide la naturaleza de las acciones de desarrollo personal que deberá asumir después del proceso de reflexión sobre su actividad.

La autoevaluación aporta un tipo de conocimiento único en el contexto de la evaluación comprensiva, que no debemos desaprovechar y cuya importancia responde a las siguientes razones (McColskey y Egelson, 1993; Airasian y Gullickson, 1995,96):

- Es un acto de responsabilidad profesional.
- Es una actividad que permite focalizar el desarrollo profesional y la mejora al nivel del aula y del centro, ámbitos donde realmente el profesor es un gran experto.
- Supone un reconocimiento que el cambio en la organización exige también cambios en los procesos personales
- Implica también el reconocimiento que el profesor mantiene el control sobre su propia práctica.
- Facilita que el profesor sea consciente respecto de las fortalezas y debilidades de su propia práctica.
- Motiva actitudes proactivas de cambio y de desarrollo profesional.
- Trata a los profesores como profesionales con capacidad de automejora.
- Favorece la interacción y el debate en el marco escolar, acerca de la calidad de la docencia.

Finalizamos este apartado señalando, que posiblemente una forma de potenciar la calidad de los autoinformes es que estos se realicen después de que el profesor ha analizado información evaluativa sobre su actividad proveniente de otras fuentes (como son las que hasta ahora hemos analizado). La información previa ayuda a centrar la autorrefelexión y favorece la profundidad de los propios juicios. Independientemente responde a la idea que mantenemos que los sistemas evaluativos sobre la docencia se fundamentan sobre el análisis complementario e interactivo de fuentes documentales múltiples y diversas en su naturaleza.

El portafolios

El portafolios del profesor se ha convertido en los últimos años en uno de los instrumentos más populares y extendidos en los Estados Unidos, tanto para la evaluación como para el desarrollo profesional del profesorado. Su uso ha sido adoptado en numerosos ámbitos incluido el universitario donde se ha aplicado fundamentalmente en la formación y desarrollo profesional de profesores noveles (Borko, Michaelec y Timmons, 1996). También ha demostrado su utilidad en el reconocimiento de actividad docente de excelencia a nivel de distrito educativo en Estados Unidos (Wolf, Lichtenstein, Bartlett y Hartman, 1996) y en procesos de renovación de licencias tanto a nivel estatal como nacional en el mismo país.

En su forma más elemental el portafolio en esencia no es sino una simple colección de informaciones acerca de la práctica docente del profesor. El problema que puede tener es que sus contenidos, que pueden ser tan variados como fotografías de la vida de la escuela, notas escritas de los padres dirigidas al profesor, etc. etc., no estén

relacionados con la filosofía docente personal del profesor o con el plan de calidad docente de la escuela, en otras palabras que carezca de significación para orientar el proceso de mejora.

Para evitarlo el portafolios debería contemplar los siguientes aspectos (Schulman 1992):

- El portafolio debería estructurarse alrededor de contenidos profesionales y de los objetivos institucionales del centro.
- Debería contener ejemplos cuidadosamente seleccionados del trabajo del profesor y de los alumnos de forma que ilustrara adecuadamente hechos o situaciones clave de la práctica docente.
- Los contenidos del portafolio deberían ir acompañados de comentarios escritos, que explicaran e interpretarán el significado e importancia de los mismos.
- El portafolio tendría que constituir una colección de experiencias vividas que sirvieran de base para discutir con los colegas elementos de desarrollo profesional.

En general se admiten tres grandes propósitos para el portafolios (Wolf, Lichtenstein y Stevenson, 1997):

- Dirigido a cumplimentar requerimientos evaluativos.
- Para orientar el desarrollo profesional.
- Como ayuda en la búsqueda de empleo.

Desde nuestro insistente punto de vista del modelo comprensivo, la función básica del portafolios está en la evaluación. El portafolios ilustra con experiencias significativas y representativas la información evaluativa que poseemos y que se requiere para llevar a buen término la evaluación sistemática de la docencia. Obviamente y previa reflexión y en el contexto de mejora de la docencia la información contenida en el portafolio ayuda en la orientación del desarrollo profesional del profesor.

En el portafolios del profesor deben incluirse todos aquellos materiales que puedan en algún momento incrementar, vía ilustración, nuestro conocimiento evaluativo de la actividad docente del profesor.

Justamente el carácter adicional que aporta, es incluir en nuestra reflexión muestras de ejemplos escogidos de la actividad profesional, que en sí mismos no constituyen necesariamente información objeto de evaluación, pero que su valor añadido consiste justamente en arrojar luz a otras informaciones que si lo son.

b. Emisión de los juicios de valor respecto de la docencia

Bajo el término docencia existe un constructo muy amplio y complejo, pero llegados a esta fase del proceso evaluativo no tendremos más remedio que enunciar como vehicular su concreción. Si hemos seguido cada uno de los pasos hasta aquí explicitados: análisis de las necesidades institucionales y su compatibilización con las expectativas profesionales del profesorado, determinación de los objetivos institucionales, desarrollo de criterios e indicadores, establecimiento de estandars, tenemos todos los elementos básicos que nos delimitan operativamente cual es el modelo de calidad que se pretende en ese centro para la docencia.

El modelo de calidad de la docencia es el que nos va a orientar en la procedimentalización del diseño evaluativo. Nos va a indicar que elementos de la docencia son los más significativos para un momento determinado y por tanto estaremos en condiciones de determinar que tipo de información deberemos recoger, a través de que instrumentos, con que secuenciación y que tipo de análisis habrá que aplicar. Todos ellos constituyen los elementos clave del diseño u operativización del procedimiento evaluativo.

No hemos de minusvalorar los aspectos procedimentales, parodiando la frase de que democracia es procedimiento, podemos afirmar que sin procedimiento no hay evaluación.

En muchísimas ocasiones agotamos la actividad evaluadora en los procesos de recogida de la información. Documentar la actividad evaluadora de manera múltiple y enriquecida, tal como hemos indicado, es un paso esencial, pero no suficiente para alcanzar la sistemacidad y la comprensividad en la evaluación.

Es preciso que dicha recogida de información esté orientada adecuadamente, que las condiciones previas facilitadoras de la evaluación se hayan cumplido, que los nexos relacionales de la evaluación con los procesos de mejora y de desarrollo profesional se cumplan, etc. etc. y todo esto nunca será posible si cada una de las acciones no están convenientemente integradas en el procedimiento evaluativo.

En definitiva la línea argumental que va desde tener exquisito cuidado con que se cumplan las condiciones de producción de la evaluación, que pasa por asegurar la calidad técnica en los procesos de recogida de información de forma que que garanticen su sustantividad y su riqueza y variedad de puntos de vista, que continua controlando el cumplimiento de los aspectos relacionales entre la información analizada y los procesos de mejora de la docencia en la institución y el desarrollo profesional de los implicados, que se desarrolla bajo un cuidado absoluto del derecho a la confidencialidad y de los derechos legales que asisten a los implicados constituyen la línea maestra que vertebra una evaluación que aspire a adquirir las características de sistemática y comprensiva.

Sólo a partir de una adecuada explicitación de cada uno de los momentos del procedimiento a efectuar es posible sistematizar el proceso entendiéndolo como un continuo en el tiempo y recorrer cada uno de los aspectos que le permite adquirir el carácter de comprensivo, ya que atiende simultáneamente a todas las audiencias en el contexto de las necesidades institucionales y profesionales expresadas en el modelo de calidad de la docencia.

c. Uso intensivo y extensivo de la información evaluativa: la derivación de las propuestas de mejora

Una de las características que mejor caracteriza las sociedades modernas, está en su evolución hacia nuevas formas de funcionamiento basadas en el uso intensivo y extensivo de la información (Tedesco, 2000). Con esta afirmación previa, queremos dar especial relevancia a una realidad que consideramos fundamental en el marco de la evaluación actual, el hecho que la derivación de las propuestas de mejora deberá asentarse sobre un análisis intensivo de la información y en su uso extensivo.

La información evaluativa ha de ser analizada en profundidad utilizando para ello todos los métodos que la inteligencia y las nuevas tecnologías sean capaces de aportar. Sus resultados deberán ser integrados en todos los ámbitos desde los que se pueda incidir en la mejora global de la calidad de la docencia.

Para conseguirlo, en el análisis de los puntos fuertes y débiles y en las propuestas de mejora deberemos atender a una doble orientación:

Por un lado el análisis para poder alcanzar el adecuado nivel de profundidad, relevancia e intensidad deberá ser orientado por el modelo de calidad de la docencia y obviamente hacer uso de los procedimientos técnicamente más avanzados y correctos.

Por otro, si queremos que los resultados incidan extensivamente, deberemos reclamar la existencia de una plan estratégico del centro donde integrar la información evaluativa con otras de distinta procedencia, pero igualmente relevantes, y desde él proyectar cualquier propuesta de mejora.

La toma de decisiones no depende únicamente de la información evaluativa, sino que su vehiculación exige integrar la información evaluativa con otras, analizarlas conjuntamente de forma realmente comprensiva e intensiva de forma que la generación de propuestas se asiente realmente en el dominio de la realidad.

El plan estratégico es el único espacio que facilita toda esta compleja red de interacciones y todo centro debería tenerlo de forma que se racionalicen y se fijen las prioridades donde inscribir cualquier propuesta de desarrollo y de cambio, entre las que evidentemente se encuentra los planes de mejora para la docencia.

Pensemos que antes de tomar cualquier decisión debe analizarse en profundidad la capacidad total del sistema, donde entre otros aspectos se analiza el potencial de cambio del centro, los recursos con que se cuenta, las necesidades, las prioridades y con todo ello se establecen las líneas maestras sobre las que discurrirá su crecimiento y desarrollo durante un período determinado y es en ese marco donde se inscribe la información evaluativa que evidentemete modifica el propio plan estratégico al incorporar sus aportaciones pero que al tiempo dimensiona y direcciona toda las tomas de decisiones ligadas con la evaluación, el desarrollo y la mejora.

Los planes estratégicos constituyen en la actualidad uno de los instrumentos más utilizados para dinamizar el desarrollo de las instituciones. Constituyen la corriente principal de cualquier organización alrededor de la que se organizan todas las iniciativas, dotándolas de sentido global y de significado, al tiempo que racionalizan los procesos de priorización y de secuenciación de las acciones a ejecutar. Facilitan las tomas de decisiones realistas e integradas en una misión común que gracias a que está explicitada puede ser debatida, mejorada y compartida.

Un modelo comprensivo de evaluación sistemática de la docencia encuentra en este espacio el contexto donde debe situarse todas las propuestas de mejora, agrupadas y ordenadas. El plan estratégico de la organización, es el único ámbito donde es posible asegurar mínimamente la racionalidad y el realismo de la acción evaluadora. Al tiempo todo plan estratégico lleva implícito su propio plan de seguimiento lo que permite cerrar eficazmente el círculo que enlaza la evaluación con la mejora.

4. La fase metaevaluativa

Desde que en 1975 se creó el «Joint Committee on Standars for Educational Evaluation» ha ido generando diversas recomendaciones para conducir procesos metaevaluativos. En lo que hace referencia al personal ha emitido dos conjuntos de normas, una el 1988 y otra en el 1994.

Según sus propias palabras se conceptualiza el término estandar como «un principio comunmente aceptado para ser usado en evaluar la calidad en educación» (Sanders 1994). Sin embargo desde nuestro punto de vista, las recomendaciones tal como ellos las emiten, las concebimos más como medios para desarrollar la calidad de las prácticas evaluativas.

Los estandars para la evaluación del personal de 1994, vienen agrupados (como en la anterior ocasión) en cuatro categorías que tratan de reflejar los atributos que debe reunir una evaluación del profesorado correcta: propiedad, utilidad, factibilidad y precisión.

Propiedad

El estandar de propiedad establece la expectativa de que las evaluaciones que afecten al profesorado deben ser conducidas legal y éticamente. Aparte deben suponer algún tipo de mejora para aquellos que van a ser evaluados y a los que sirven. Se enfatiza en este estandar los siguientes aspectos:

- La orientación hacia el servicio: La evaluación del profesorado debe promover principios educativos correctos, la protección de los derechos del estudiante y de la comunidad y la satisfacción de las necesidades del sistema educativo. Los estandares específicos para la enseñanza deben ser formulados de forma que refuercen consistentemente a todos los profesores.
- Carácter formal: Las políticas, los procedimientos y los criterios que guían la evaluación del profesorado deben ser establecidos de manera consistente, equitativa y legal.
- Conflicto de intereses: La evaluación debe estar protegida de procesos inapropiados que puedan sirvan a intereses personales discriminatorios. Su orientación debe ser el servicio al centro y a objetivos profesionales compartidos.
- Acceso a los informes personales: La evaluación del profesorado tiene carácter confidencial y su acceso debe estar perfectamente controlado.
- Interacción con los evaluados: Los profesores evaluados debereán ser tratados de manera profesional, considerada y cortés.

Utilidad

La evaluación del profesorado ha de ser informativa, oportuna e influyente. En este sentido se cuidarán específicamente los siguientes elementos:

- Orientación constructiva: La evaluación del profesorado deberá incidir en la mejora de la calidad del centro, en el desarrollo de sus recursos humanos y asistir a los profesores en la mejora del cumplimiento de sus responsabilidades.

- Usos definidos: Los usuarios y los usos derivados de la evaluación del profesorado deberán ser identificados desde el principio del proceso evaluativo.
- Credibilidad del evaluador: La evaluación del profesorado deberá ser dirigida y conducida por personal cualificado, entrenado, autorizado, soportado y monitorizado.
- Informe funcional: El informe producto de la evaluación del profesorado deberá ser claro, oportuno, preciso y pertinente. Debe en definitiva tener valor práctico para los usuarios.
- Seguimiento e impacto: No basta con meramente informar, una evaluación de calidad proporcionará ayuda a los usuarios para que tomen las acciones pertinentes derivadas de la información evaluativa.

Factibilidad

La evaluación del profesorado deberá ser de fácil implementación, eficiente, adecuada apoyada y viable desde la perspectiva de la comunidad. Entre sus características tendrá que observar:

- Procedimientos prácticos: Los procedimientos para obtener la información deben ser justo los adecuados y todo el proceso debe tener el mínimo costo y provocar las mínimas interrupciones en la práctica diaria.
- Viabilidad política: Las políticas y los procedimientos a establecer procurarán ser aceptables para todos los implicados.
- Viabilidad fiscal: El tiempo y los recursos deberán ser los adecuados, evitando peligro de sobredimensionarlos.

Precisión

Este estandar hace referencia a la fidelidad del proceso. Así se exige que las conclusiones deberán estar absolutamente fundamentadas en las evidencias analizadas. Pensemos que al ser una evaluación que su objeto directo son personas habrá que exigir un esmerada atención a los detalles y a la precisión en las afirmaciones. Entre sus características señalamos:

- Fuciones y responsabilidades bien definidas: Todas las partes involucradas en este tipo de evaluación, deberán tener homogeneamente asumidos y entendidos cuáles son las responsabilidades o funciones (y su significado) que forman parte de los deberes del profesor y que van a constituir objeto de la evaluación.
- Ambiente de trabajo: El contexto donde trabaja el profesor es evidente que puede afectar a su trabajo, por tanto la evaluación deberá tener en consideración las posibles influencias derivadas del mismo.
- Procedimientos de documentación: La evaluación deberá estar adecuadamente documentada, de forma que se pueda acudir a la misma para clarificar los extremos evaluativos.
- Validez y fiabilidad de los instrumentos, estrategias y procedimientos usados en la recogida de información.

- Control sistemático de los datos: Hay que garantizar que la información recogida para desarrollar los procesos evaluativos, será guardada de manera segura y que será procesada cuidadosamente de forma que no se introduzcan errores durante el proceso de manipulación de los datos.
- Control de los sesgos: La evaluación deberá ser limpia e imparcial para todos los profesores. Se debe asegurar la protección contra toda información afectada por cualquier tipo de parcialidad.
- Control sistemático de la evaluación: Será preciso revisar periódica y cuidadosamente los sistemas evaluativos de forma que se vayan introduciendo de forma continua las pertinentes mejoras.

EVALUACIÓN Y DESARROLLO PROFESIONAL

Entramos ya en la recta final de nuestro trabajo y aunque sea muy brevemente queremos aportar alguna reflexión sobre el tema del desarrollo profesional que evidentemente constituye una de las consecuencias derivada de los procesos de evaluación de la docencia.

Trataremos de exponer algunos cambios que se están produciendo en cuanto a su concepción. Así en los estándares establecidos por el National Staff Development Council (1995a, 1995b, 1995c) se introducen dos ideas que consideramos de máximo interés y que pasamos a comentar a continuación.

En primer lugar hay que romper con la imagen que cuando hablamos de desarrollo profesional se está tratando de algo que afecta únicamente a los profesores. En la actualidad se entiende como un proceso de crecimiento que afecta a toda la institución, aún es más a toda la comunidad educativa, lo cual compromete a todas las personas que tengan algún impacto en la docencia, (p.e.: se pueden incluir hasta a los padres), como también a sus aspectos organizativos e incluso a los recursos que se aplican.

Estamos frente a un modelo que entiende que el desarrollo profesional es un compromiso global del centro y que todos aquellos elementos que intervienen favoreciendo la mejora profesional de las actuaciones en el centro son objeto de atención en las acciones de desarrollo profesional.

En otra línea está la segunda aportación del National Staff Development Council, por la que se amplía el campo de actuaciones de mejora para el desarrollo profesional dirigidas a los profesores. Según esta segunda aportación, no deberemos circunscribir las intervenciones para el desarrollo profesional a la formación presencial personal o de grupo del profesorado, sino que en el desarrollo profesional se avanza y se realiza fundamentalmente a través de la propia práctica profesional y afectando a todas sus partes.

Entendemos que estas aportaciones completan mejor nuestras propuestas de modelo comprensivo para la evaluación sistemática de la docencia.

Así y como resumen y epílogo, sólo nos queda recordar que iniciábamos nuestro recorrido insistiendo que la evalución del profesorado cobra significado cuando se integra en el concepto de evaluación de la docencia, entendiendo la docencia como

responsabilidad común de la institución especialmente de sus equipos docentes, a su vez la docencia responde a las directrices del modelo de calidad del centro.

Seguíamos con que la recolección de información para la evaluación si quiere atender a tan vasto dominio deberá conjugar el uso de múltiple documentación para la captura de toda la gama posible de matices implícitos en la información evaluativa con la de otras informaciones de procedencias diversas. La recolección y uso intensivo de la información será preciso para entender y atender la realidad educativa. Esta información deberá integrarse en la corriente común de los planes estratégicos si queremos procedimentalizar no tan sólo la acción evaluativa sino también la construcción y seguimiento de planes de mejora realistas.

Finalmente concluíamos que el desarrollo profesional entendido como una de las acciones derivadas de la evaluación deberá superar anteriores concepciones y entenderse como una acción que compromete a toda la institución y que se ejecuta fundamentalmente en el espacio profesional por ella enmarcado.

BIBLIOGRAFÍA

AIRASIAN, P.W. y GULLIKSON, A.

(1995) *Teacher self-evaluation tool kit.* Raport presentado a la Office of Research, Office of educational Research and Improvement, U.S. Department of Education, Washington, D.C.

(1996) *Teacher self-evaluation tool kit.* Thousand Oaks, Ca.: Corwin.

BECHER, R.M. (1984). *Parent involvement: A review of researchand principles of succesful practice.* (Report No. PS 014 563). Urbana, Il.: National Institute of Education (ERIC Document Reproduction Service No. ED 247 032).

BERK, R.A. (1988) Fifty reasons why student achievement gain does not mean teacher efectiveness. *Journal of Personnel Evaluation in Education,* 1, 345-363.

BORKO, H., MICHAELEC, P. y TIMMONS, M. (1996). *Student teaching portfolios: A tool for promoting reflective practice.* Washington D.C.: Author.

CASTETTER, W.B. (1981) *The personnel function in educational administration (3ª. ed)* New York: Macmillan.

CONNELLAN, T.K. (1978) *How to improve human performance: Behaviorism in business and industry.* New York: Harper and Row.

CONLEY, D.T. (1987) Critical attributes of effective evaluation systems *Educational Leadership,* 44 (7), 60-64.

DELORS, J. (1996) Los cuatro pilares de la Educación en *La Educación encierra un tesoro.* París: UNESCO.

DUKE, D.L. (1990) Developing teacher evaluation systems that promote professional growth. *Journal of Personnel Evaluation in Education,* 4, 131-144.

ELLET, C.D. (1987) Emerging teacher performance assessment practices: Implications for the instructional supervision role of school principals en W. Greenfield (ed.) *Instructional leadership: Concept and controversies* (pp. 302-307) Boston: Allyn y Bacon.

EPSTEIN, J. (1985). A question of merit: Principals' and parents' evaluation of teachers. *Educational Researcher,* 14(7), 3-10.

EVERSTON, C.M. y HOLLEY, F.M. (1981) Classroom observation en J. Millman (ed.) *Handbook of teacher evaluation,* pp. 90-109. Beverly Hills, Ca.: Sage.

FAUCETTE, S.L., BALL, D.M. y OSTRANDER, L.R. (1995). *Clients satisfaction surveys.* Virginia Beach,Va.: Virginia Beach City Public Schools Press.

FOLLMAN, J. (1992). Secondary school students' ratings of teacher effectiveness. *High School Journal,* 75, 168-178.

FRASE, L. y STRESHLY, W. (1994) Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education,* 1, 47-57.

FULLAN, M.G. (1991) *The new meaning of educatoinal change.* New York: Teachers College Press.

GLATTHORN, A. (1984) *Differentiated supervision.* Alexandria, VA.: Association for Supervision and Currículum Development.

GOODALE, J.G. (1992) Improving performance appraisal *Business Quarterely,* 51 (2) 65-70.

HAAK, R.A., KLEIBER, D. y PECK, R. (1972) *Student evaluation of teacher instrument, II.* Austin: Texas University, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 080 574).

HARRIS, B.M. (1987) Resolving old dilemmas in diagnostic evaluation en *Educational Leadership,* 44 (7), 46-49.

HELM, V.M. (1994, April) *The professional support personnel evaluation model: The use of multiple data sources in personnel evaluation.* Ponencia presentada al encuentro anual de la American Educational Research Association, New Orleans.

HUTCHMACHER, W. (1999) L'avaluació en la transformació de les modalitats de govern dels sistemes educatius. Tendències europees en *Avaluació i Educació,* pp. 15-34. Barcelona: Generalitat de Catalunya. Consell Escolar de Catalunya.

JOINT COMMITTEE ON STANDARS FOR EDUCATIONAL EVALUATION (1988) *The personnel evaluations standars: How to assess systems of evaluating educators.* Newbury Park, C.A.: Sage.

MAGER, G.M. (1980). The conditions which influence teachers in initiating contacts with parents. *Journal of Educational Research,* 73, 276-282.

MARCH, J.G. y SIMON, H.A. (1967). *Organitzations.* New York: John Wiley.

MARK, M.M., y SHOTLAND, R.L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review,*9, 605-626.

MATEO, J., ESCUDERO, T., MIGUEL, M., MORA, J.G. y RODRÍGUEZ, S.

(1996) *La evaluación del profesorado de secundària. La opinión de los profesores.* Barcelona: CEDECS.

(1999) La evaluación del prfesorado de secundária en Cerdán, J. y Grañeras, M. *La investigación sobre profesorado II,* 1993-97, pp. 339-359.

McCOLSKEY, W. y EGELSON, P. (1993). *Designing teacher evaluation systems that support professional growth.* Greensboro, N.C.: South-estearn Regional Vision for Education.

MCGREAL, T.L. (1988) Evaluation for enhancing instruction. Linking teacher evaluation and staff development en S.J. Stanley y W.J. Popham (eds.), *Teacher evaluation: Six prescriptions for success* (pp. 1-29), Alexandria, VA.: Association for Supervision and Curriculum Development.

MC.LAUGHLIN, M.W. y PFEIFER, R.S. (1988) *Teacher evaluation: Improvement, accountability and effective learning.* New York: Teachers College Press.

MILLMAN, J. y SYKES, G. (1992) *The assessment of teaching based on evidence of student learning: An analysis* (Research Monograph No.2). Washington, D.C.: National Board for Professional Teaching Standars.

NATIONAL STAFF DEVELOPMENT COUNCIL.

(1995 a) inCooperation with the National Association of Elementary School Principals. *Standars for staff development* (Elementary school ed.). Oxford, Oh.: Author.

(1995 b) in Cooperation with the National Association of Secondary School Principals. *Standars for staff development* (High school ed.). Oxford, Oh.: Author.

(1995 c) *Standars for staff development* (Middle school ed.). Oxford, Oh.: Author.

OLSEN, L.D. y BENETT, A.C. (1980). Performance appraisal: Management technique as social process? En D.S. Beach (ed.) *Managing people at work: Readings in personnel (3ª. Ed.)* New York: Macmillan.

PATTON, M. Q. (1986) *Utilization-focused evaluation (2ª. ed.)* Beverly Hills, CA.: Sage.

PHI DELTA KAPPA NATIONAL STUDY COMMITTEE ON EVALUATION (1971) *Educational evaluation and decision making.* Itaca, Il.: F.E. Peacock.

ROOT, D. y OVERLY, D. (1990) successful teacher evaluation: Key elements for success. *NASSP Bulletin,* 74(527), 34-38.

SANDERS, J.R. (1994) The process of developing national standars that meet ANSI guidelines. *Journal of Experimental Education,* 63 (1), 5-12.

SCRIVEN, M.

(1981) Summative teacher evaluation en J. Millman (ed.) *Handbook of teacher evaluation pp. 244-271.* Beverly Hills, Ca.: Sage. (1988). «Duties-based teacher evaluation. *Journal of Personnel Evaluation in Education,* 8, 151-184.

(1991) *Duties of teacher* (TEMP A memo) Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

SCHULMAN, L. (1992) *Portfolios for teacher education: A component of reflective teacher education.* Ponencia presentada al encuentro anual de la American Educational Research Association, San Francisco. Ca.

SMYLIE, M.A. (1994) «Redesigning teachers' work: Connections to the classroom» en L. Darling-Hammond (ed.), *Review of research in education* (pp. 129-178). Washington, D.C.: American Educational Research Association.

STRONGE, J.H. (1997) *Evaluating Teaching. A guide to current thinking and best practice.* Thousand Oaks, Cla.: Corwin Press, Inc.

STRONGE, J.H. y OSTRANCHER, L.P. (1997) «Client Surveys in Teacher Evaluation en Stronge, J. H. (ed.) *Evaluating Teaching. A guide to current thinking and best practice.* Thousands Oaks, Cla.: Corwin Press. Inc.

STRONGE, J.H., HELM, V.M., y TUCKER, P.D. (1995) *Evaluation handbook for professional support personnel.* Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

TEDESCO, J.C. (2000) Educación y sociedad el conocimiento *Cuadernos de Pedagogía,* 288, pp. 83-86.

WEBSTER, W.J. (1994). The connection between personnel evaluation and school evaluation. En A.A., McConney (ed.), *Toward a unified model: The foundations of educational personnel evaluation* (pp. 93-117). Kalamazoo: Western Michigan University Center for Research on Educational Accountabilioty and Teacher Evaluation.

WHEELER, P.H. y HAERTEL, G.D. (1993), *Resource hanbook on performance assessment and measurement: A tool for students practitioners, and policymakers.* Chico, Ca.: Owl. (ERIC Document Reproduction Service No. ED 367 686).

WOLF, K., LICHTENSTEIN, G., BARTLETT, E. y HARTMANN, D. (1996) Professional development and teaching portfolios: The Douglas County School District Outstanding Teacher Program. *Journal of Personnel Evaluation in Education,* 10, 279-286.

WOLF, K., LICHTENSTEIN, G. y STEVENSON, C. (1997) Portafolios in Teacher Evaluationen J.H. Stronge, *Evaluating Teaching. A guide to Current Thinking and best Practice.* (pp. 193-213) Thousands Oaks, Ca.: Corwin Press, Inc.

WOOD, C. (1992) Toward more effective techer evaluation: Lessons from naturalistic inquiry. *NASSP Bulletin,* 76 (542), 52-59.

TRABAJOS DE INVESTIGACIÓN

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 37-45

LA OPINIÓN DE LOS ESTUDIANTES UNIVERSITARIOS SOBRE EL MÉTODO DOCENTE DE LAS FACULTADES DE CIENCIAS

Francisco José Sáez
Universidad del País Vasco / Euskal Herriko Unibertsitatea*

RESUMEN

El objetivo de este estudio era conocer la opinión de los alumnos sobre el método docente utilizado en una Facultad de Ciencias. Se realizó una encuesta a los alumnos con preguntas sobre: 1) método docente preferido y 2) método de estudio personal (hábitos y técnicas de estudio). Algunos alumnos se sometieron voluntariamente a un método docente basado en el trabajo personal diario y en la evaluación continua, y contestaron otra encuesta para conocer su opinión sobre: 1) el sistema de aprendizaje preferido y 2) el mecanismo de evaluación preferido. Los resultados indican que los alumnos prefieren unas clases explicativas y participativas a otras meramente expositivas (magistrales), y que ellos consideran que deben potenciarse las técnicas de comprensión frente a las de memorización. El análisis estadístico demuestra que pueden distinguirse dos poblaciones estudiantiles con diferente concepción de la actividad docente y discente, que coinciden con los dos grupos de alumnos estudiados.

ABSTRACT

The objective of this study was to know the opinion of the pupils about the university educational activity in a Sciences Faculty. A survey to the pupils was accomplished with

* Departamento de Biología Celular y Ciencias Morfológicas. Facultad de Medicina y Odontología. 48940 Leioa (Vizcaya). E-mail: gcpsacrf@lg.ehu.es

BEST COPY AVAILABLE

questions about 1) educational method preferred, and 2) method of study (students' habits and techniques). Some pupils were submitted to an educational method based on the daily personal work and the continuous evaluation, and they answered other survey to know their opinion about: 1) preferred learning system, and 2) preferred mechanism of evaluation. The results indicate that the students prefer an explanatory classes to other merely magisterial lectures and consider that it must be prior the comprehension as compared to the memorization. The statistic analysis demonstrates that two student populations with a different conception of the educational activity, which coincide with the two groups of studied pupils, can be distinguished.

1. INTRODUCCIÓN

Actualmente, el método docente más utilizado en las Facultades y Escuelas Universitarias de ciencias experimentales es, probablemente, la llamada lección magistral. La lección magistral es, sin duda, una herramienta muy útil, por cuanto permite la presentación al alumno de mucha información, de forma seleccionada y resumida. Para el alumno, supone una gran ayuda que el profesor condense los aspectos más importantes de la asignatura dado que, en ocasiones, ésta está dispersa en múltiples fuentes de información. Sin embargo, la lección magistral tiene una serie de desventajas. Así, la gran cantidad de información existente en muchas materias, y el cada vez menor tiempo lectivo disponible para su exposición en clase, hacen que el profesor se limite a exponer o dictar una serie de conocimientos, lo que implica que el alumno se convierta en una máquina de copiar, ya que no hay tiempo en clase para la reflexión o la discusión. Como resultado de todo ello, en nuestras aulas se aprecia una tendencia cada vez mayor, por parte de los estudiantes, a limitar el hábito de estudio a las fechas próximas a los exámenes finales y a reducir los contenidos a aquello que se ha estudiado en clase, es decir, a ser meros repetidores de aquello que han copiado.

En este contexto, puede ser interesante conocer la opinión de los estudiantes sobre el método docente. Aunque pueda ser discutible el valor que posee la opinión de los alumnos para valorar un método docente, no es la primera vez que se opta por este abordaje (Mallado Rodríguez y Pérez López 1995). Además, en muchas universidades se están realizando procesos de evaluación del profesorado universitario que, generalmente, incluyen conocer la opinión de los alumnos sobre los profesores (Escotet 1984, Sobrado Fernández 1991).

Por ello, el autor de este trabajo decidió conocer cuál era la opinión de sus alumnos sobre el método docente, para lo cual realizó una encuesta entre los estudiantes de la asignatura Citología e Histología de la Facultad de Ciencias de la UPV/EHU. Algunos de estos alumnos se sometieron voluntariamente durante el curso a un método docente basado en el trabajo personal diario y en la evaluación continua.

Los resultados indican un claro rechazo a un método docente, ampliamente utilizado en las facultades de ciencias experimentales, basado exclusivamente en la exposición oral del tema, así como un grado aceptable de satisfacción por el estímulo al estudio diario. Además, se aprecia la coexistencia de dos subpoblaciones de alumnos: una habituada al sistema actual, con menor tendencia a demandar otro tipo de activi-

dades, y que coincide con los alumnos que no se sometieron a la evaluación continua; y otro grupo formado por los alumnos que se presentaron voluntarios a la evaluación continua, que muestran claramente una mayor predisposición a un método docente más participativo y mayor rechazo de la utilización exclusiva de la lección magistral.

2. METODOLOGÍA

Noventa y cuatro alumnos de la asignatura «Citología e Histología» de 2º curso de la licenciatura en Ciencias Biológicas de la Facultad de Ciencias de la UPV/EHU, contestaron un cuestionario (Anexo I, cuestionario A) sobre la actividad educativa universitaria. En dicho cuestionario, el alumno podía responder a cada pregunta según el grado de acuerdo o asentimiento a la misma, en una escala del 1 al 5, donde 1 significaba total desacuerdo y el 5 asentimiento total. Además, se incluyó un apartado de observaciones donde el alumno pudiera reflejar de forma más concreta su opinión personal. Este cuestionario incluía preguntas que pretendían discernir su opinión sobre las siguientes variables:

1. Método de estudio, en concreto sobre el hábito de estudio diario (preguntas 1º y 2º) y sobre las técnicas de estudio (ítems 8º, 9º y 10º); y
2. Método docente preferido (preguntas 3º, 4º, 5º, 6º y 7º).

En los ítems número 1º, 9º y 10º, relativos al método de estudio, las respuestas de asentimiento (valor próximo al 5) indican una predilección por el estudio diario y el estudio reflexivo, no de memorización, al contrario que en las cuestiones número 2º y 8º, donde dicha preferencia queda indicada con respuestas de desacuerdo (valores próximos al 1). En las cuestiones referidas al método docente, la preferencia por unas clases más participativas y un rechazo a la lección magistral se pone de manifiesto con respuestas próximas al 1 en las preguntas 3º, 4º y 5º, y con respuestas próximas al 5 en los ítems 6º y 7º.

De los 94 alumnos encuestados, 56 se sometieron voluntariamente a un método docente basado en el trabajo diario y en la evaluación continua, y los otros 38 fueron evaluados mediante un examen tradicional. El método consistía en la preparación de los temas por los alumnos y una breve exposición ante sus compañeros. Además, algunos alumnos, de forma voluntaria, prepararon seminarios sobre temas no incluidos en el programa, que expusieron en clase. Al final del curso, se pidió a los 56 alumnos que colaboraron en el sistema de evaluación continua, que contestaran otro cuestionario (Anexo II, cuestionario B) de similares características que el cuestionario A, con preguntas sobre:

(a) El sistema aprendizaje preferido (números 1º, 2º, 3º y 4º).
(b) El mecanismo de evaluación preferido (en general todas las preguntas, pero especialmente la número 5º).

En este cuestionario, la preferencia por el trabajo y el estudio diarios y la evaluación continua se corresponde con valores próximos al 5 en las preguntas 1º, 2º y 4º, y por respuestas próximas al 1 en las cuestiones número 3º y 5º, mientras que el mayor grado de acuerdo con el método de evaluación tradicional se pone de manifiesto por las respuestas tendentes a los valores contrarios en cada ítem.

Se estudiaron las respuestas de los 94 alumnos encuestados y, además, se analizaron los resultados dividiendo a los alumnos en dos subpoblaciones: los alumnos que se examinaron de la forma tradicional (grupo I, 38 alumnos), y los que se sometieron a evaluación continua (grupo II, 56 alumnos). Se analizaron y compararon los resultados de ambos grupos con el fin de ver si había diferencias significativas entre ambas subpoblaciones. Se calculó la media aritmética de los valores obtenidos para cada pregunta, y se compararon ambos grupos mediante el test de la «t» de Student.

3. RESULTADOS Y DISCUSIÓN

Los resultados del cuestionario A se muestran en la tabla 1. Como puede verse, se aprecian diferencias significativas entre ambas poblaciones en las respuestas a las preguntas 3º, 5º, 7º y 10º. Los resultados del cuestionario B se muestran en la tabla 2.

El método docente más extendido en la enseñanza universitaria en nuestro país, tanto en las llamadas enseñanzas experimentales como en gran parte en las humanísticas, es sin duda la llamada «lección magistral» (Mallado Rodríguez y Pérez López, 1995). El autor del presente trabajo se plantea investigar la mejora de su método docente en el contexto de su actividad universitaria. Como primera aproximación, en este trabajo se trata de conocer la opinión de los alumnos sobre el método docente y sus hábitos de estudio.

Los resultados muestran un elevado rechazo al método docente utilizado con mayor frecuencia en las facultades de ciencias experimentales (especialmente en la Licenciatura en Biología), que podríamos asegurar que se basa fundamentalmente en la exposición de datos por parte del profesor, y en la toma de apuntes (a veces al dictado), por el alumno. Esto suele ir unido a una carencia de estímulo a la reflexión, a la discusión, y a la ampliación de conocimientos por el estudiante y, además, a una evaluación final consistente en la repetición de lo que el profesor «explicó» (y sólo eso).

En efecto, si analizamos las respuestas a las preguntas 3º a 6º, vemos un rechazo manifiesto a la lección magistral (cuestión 5º), si bien no hay un rechazo total al hecho de que el profesor dicte apuntes (cuestión 6º). Esta aparente contradicción refleja, probablemente, que el método actual es cómodo y efectivo para el alumno, puesto que simplifica la labor de estudio con vistas al mero hecho de superar un examen. Por lo tanto, es práctico a la hora de aprobar. Además, los alumnos poseen pocas referencias claras de otros métodos docentes (Mallado Rodríguez y Pérez López, 1995). Sin embargo, el alumno tiene conciencia de que este método, cómodo y práctico, no le permite en realidad un aprendizaje efectivo, real, de la materia objeto de estudio.

Estas conclusiones se ven reforzadas si analizamos las respuestas a las preguntas 1º-2º y 8º-10º. En este caso, vemos claramente que el alumno desea un método docente basado en la reflexión y la comprensión, y no en la simple memorización. Al responder estos ítems, los alumnos se decantan claramente por la comprensión frente a la memorización, al estudio diario contra el de los últimos días, al trabajo personal contra el resumen dictado por el profesor.

El hecho de que los alumnos piensen mayoritariamente que es necesario el estudio diario (número 1), algo que muy pocos cumplen, es significativo de que el método docente basado en la lección magistral no estimula este hábito. Cabe preguntarse entonces qué hace que la lección magistral sea, no ya el método docente más extendido; sino casi el único que se practica en muchas aulas universitarias. Santos Guerra (1992) indica que la actividad de los docentes universitarios está sometida a diversos condicionantes que le hacen perpetuarse y repetirse de forma casi monótona; y plantea dos tipos de motivos:

- Motivos diacrónicos: la imagen que tenemos de profesores que, cuando eran alumnos los que ahora imparten clase, imperaba en la Universidad.
- Motivos sincrónicos: es decir, el trabajo repetitivo día a día, año tras año, en un contexto institucional y profesional poco favorable al cambio.

Otro hecho importante que destaca si analizamos detenidamente los resultados, es la coexistencia de dos subpoblaciones entre los alumnos, que coinciden con los que optaron por el método de evaluación continua y los que prefirieron el examen tradicional. Por tanto, el hecho de que algunos alumnos escogieran el método basado en el trabajo diario y otros no, es simplemente un reflejo de la coexistencia entre ellos de dos poblaciones diferenciadas: un conjunto de alumnos habituados en mayor o menor medida al estudio diario y predispuestos por tanto al trabajo intenso, que constituyen el grupo II que se sometió a la prueba; y otra población de alumnos que, por hábito, han optado por el estudio para el examen final y rechazan, de forma más o menos consciente, el trabajo diario.

La afirmación anterior se sustenta claramente mediante el análisis de las diferencias encontradas en las respuestas al cuestionario A. Así, en las respuestas a las preguntas 3º, 5º, 7º y 10º hay diferencias significativas. Concretamente, estas preguntas hacen referencia a cuál debe ser la actitud del profesor o del alumno en el aula, es decir, el método docente que debemos emplear. Los alumnos del grupo I muestran mayor tendencia hacia un profesor repetidor de datos, que no pregunte en clase, que no exija esfuerzos a diario. En cambio, el grupo II (alumnos que optaron por la evaluación continua) tiende claramente a lo contrario, es decir, a demandar un profesor que dialogue, pregunte y responda a las preguntas.

Podría pensarse que esta división entre dos poblaciones estudiantiles es artificial, ya que en el resto de las preguntas no se observan diferencias significativas. Si analizamos el contenido de dichas preguntas vemos que esto no es así. Estas cuestiones hacen referencia fundamentalmente, no a la actitud del profesor (método docente), sino a la actitud del propio alumno (método de aprendizaje). Aunque probablemente ambas poblaciones muestren hábitos de estudio diferentes, estas diferencias no se ponen de manifiesto en los resultados por varias razones. En primer lugar, aunque un alumno no tenga hábito de estudio diario, generalmente reconoce que éste es fundamental para retener conceptos. En segundo lugar, si un alumno que piensa de forma contraria es preguntado sobre el tema, raramente va a reconocer que sus hábitos de estudio son defectuosos, que sólo memoriza o que no suele entender lo que estudia. Vemos que se solapan dos factores aparentemente opuestos, el reconocimiento de lo que se hace o no se hace bien y la ocultación de este

mismo hecho. Ello hace, en resumen, que los datos con respecto a ambas poblaciones no muestren diferencias.

Por otra parte, los alumnos sometidos a evaluación continua, que respondieron a las preguntas del cuestionario B, se muestran satisfechos por el estímulo al trabajo y al estudio diarios y rechazan claramente el examen final. La crítica negativa se centró más en torno al método empleado en este caso concreto, dado que el elevado número de alumnos, la corta duración de la asignatura y el hecho de que otros alumnos no se prestaran voluntarios a este sistema, no facilitaron la labor docente.

Es significativo señalar que, en el apartado de observaciones, muchos encuestados señalaron críticas y comentarios interesantes, muchas veces en un sentido favorable al ensayo realizado y con rechazo del método tradicional. En general, los alumnos se mostraron muy críticos a la masificación en las aulas y reconocieron que este factor era el principal responsable del desarrollo indiscriminado de un sistema docente viciado, como el que actualmente sufren.

Creo que es interesante resumir algunas de las opiniones expresadas por los alumnos en este apartado. Dejando aparte aquellas observaciones que hacen referencia al método concreto seguido en la clase, o las aptitudes del profesor, y centrándonos en las que hacían referencia a consideraciones más generales, he seleccionado aquellas que me han parecido más interesantes, lo que puede conferir a esta muestra cierto grado de subjetividad:

- Si hubiera más material actualizado en la biblioteca podríamos aprender mucho más.
- Es difícil de asimilar tantos conceptos en una asignatura cuatrimestral.
- Ha faltado tiempo para realizar una buena evaluación continua.
- Era poco tiempo para poder evaluar bien.
- El profesor debe exponer los conceptos fundamentales... pero no está de más que el alumno desarrolle lo enseñado e investigue por su cuenta.
- Creo que los profesores que sólo se dedican a dar la lección magistral se deberían dedicar a otras labores.
- Ningún método es totalmente perfecto, aunque creo que sinceramente es mejor que soltar el rollo.
- Debería haber más horas de clase para dar los mismos temas... que no sean cuatrimestrales.
- Esto es lo que creo que sería mejor, pero hay que considerar lo que de verdad se podría hacer.
- Este método sería perfecto para una clase con menos alumnos...
- Creo que el método es bueno porque te obliga a estudiar día a día... yo me entero más de un tema trabajando sobre él que estudiando y repitiendo como un loro.

En resumen, podemos decir que el sistema docente basado únicamente en la lección magistral supone ciertas desventajas y conlleva un desarrollo defectuoso del proceso de aprendizaje, viciando los hábitos de estudio del alumno. Por tanto, debemos evitar caer en la tentación de utilizar la lección magistral, que evidentemente presenta muchas ventajas, como el único método docente. Se hace recomendable

simultanear este método con otras actividades que estimulen la participación y el trabajo, personal y en grupo, de los alumnos. Éstos se muestran receptivos a un cambio metodológico, pese a la habitual resistencia al cambio que siempre se aprecia. Sin embargo, el cambio metodológico debe ser realizado en condiciones de trabajo óptimas, lo que resulta difícilmente viable con la actual masificación de las aulas universitarias.

Como dice Santos Guerra (1990), «desde la interrogación sistemática sobre lo que se hace, se puede llegar a descubrir aquello que no está en la superficie, que es preciso hacer presente a través de un proceso elaborado de búsqueda». Este es un proceso paciente y constante. Por tanto, este trabajo no es más que un planteamiento de interrogantes sobre nuestra actividad docente. Para mejorar nuestra actividad debemos preguntarnos los porqués, asumir una actitud de búsqueda, de investigación sobre la práctica diaria. Es, en el fondo, la base de la investigación-acción, es decir, hacer un diagnóstico del problema (Elliot, 1990).

BIBLIOGRAFÍA

ELLIOT, J. (1990). *La investigación-acción en educación*. Madrid: Morata.

ESCOTET, M.A. (1984). *Técnicas de evaluación institucional en la educación superior*. Madrid: M.E.C.

MALLADO RODRÍGUEZ, J.A. y PÉREZ LÓPEZ, J.A. (1985). La enseñanza universitaria ante las demandas de los discentes (profesores, métodos y medios). *Revista de Enseñanza Universitaria, 9*, 61-80.

SANTOS GUERRA, M.A. (1990). *Hacer visible lo cotidiano. Teoría y práctica de la evaluación cualitativa de centros escolares*. Madrid: Akal.

SANTOS GUERRA, M.A. (1992). Una noria en el pozo de la práctica. *Revista de Enseñanza Universitaria, 2-3*, 23-32.

SOBRADO FERNÁNDEZ, L. M. (1991). Evaluación de la docencia universitaria. *Revista de Ciencias de la Educación*, 153-169.

Anexo I: Cuestionario A

1º. La base del estudio consiste en estudiar todos los días.
2º. Es suficiente con estudiar los días antes del examen.
3º. El profesor debe dictar al alumno aquello de lo que luego le va a examinar.
4º. En clase, el alumno debe copiar todo lo que dice el profesor.
5º. El tipo de clase que más me gusta es la lección magistral.
6º. La clase debería utilizarse para aclarar, discutir y/o debatir aquellos puntos más interesantes de cada tema.
7º. Me gusta que el profesor pregunte en clase.
8º. Aprendo las cosas de memoria, aunque no lo entienda.
9º. Lo más importante es comprender e interrelacionar conceptos.
10º. Es muy útil para la formación del alumno preparar e impartir seminarios (responda sólo si ha preparado algún seminario alguna vez en alguna asignatura).

Anexo II: Cuestionario B

1º. Realizando esquemas he aprendido el tema.
2º. Habría preferido que el profesor preguntara más en clase.
3º. Hemos realizado suficientes actividades en clase.
4º. Es mejor preparar el tema en casa y trabajar sobre él en clase.
5º. Prefiero estudiar para un examen final.

TABLA 1

RESULTADOS DEL CUESTIONARIO A. LOS DATOS SE MUESTRAN COMO MEDIA ARITMÉTICA DE LA OPINIÓN, REFLEJADA COMO GRADO DE ACUERDO DE LOS ALUMNOS CON LA PREGUNTA O AFIRMACIÓN PLANTEADA, DONDE 1 SIGNIFICA DESACUERDO TOTAL Y 5 ACUERDO TOTAL. ENTRE PARÉNTESIS SE MUESTRA LA VARIANZA. SE HAN SEÑALADO CON UN ASTERISCO LAS PREGUNTAS QUE PRESENTAN DIFERENCIAS SIGNIFICATIVAS

PREGUNTA	GRUPO I	GRUPO II
1º	3,9 (0,98)	4,2 (0,75)
2º	2,1 (0,82)	1,9 (0,71)
3º *	3,5 (1,33)	2,6 (1,36)
4º	2,6 (1,43)	2,2 (1,06)
5º *	2,1 (1,28)	1,6 (1,03)
6º	3,6 (1,44)	3,7 (1,31)
7º *	2,2 (1,48)	2,9 (1,25)
8º	1,9 (0,77)	1,8 (0,79)
9º	4,7 (0,31)	4,9 (0,16)
10º *	3,3 (1,87)	4,3 (1,01)

TABLA 2
RESULTADOS DEL CUESTIONARIO B. LOS RESULTADOS SE MUESTRAN COMO EN LA TABLA 1.

PREGUNTA	GRUPO II
1º	4,2 (0,57)
2º	2,9 (1,41)
3º	2,7 (1,28)
4º	4,1 (0,59)
5º	1,3 (0,43)

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 47-68

DISEÑO DE ESCALAS PARA LA EVALUACIÓN DE LA HABILIDAD MOTRIZ DE SALTO PARA LA EDUCACIÓN FÍSICA EN EDUCACIÓN PRIMARIA

Francisco Javier Castejón Oliva
Universidad Complutense de Madrid*

RESUMEN

Se ha diseñado un conjunto de pruebas con nivel de dificultad creciente, siguiendo el modelo probabilístico de Rasch, con el objetivo de evaluar la habilidad motriz de salto (contenido del área de Educación Física) del alumnado de Educación Primaria.

Los estadísticos empleados para este tipo de pruebas confirman la validez y fiabilidad de las mismas, de manera que para su aplicación se ha elegido una muestra de alumnos de las edades correspondientes a la etapa de Educación Primaria (n=1951; 1057 chicos y 894 chicas). Los resultados han permitido elaborar tareas con tres niveles de dificultad (alta, media y baja) para evaluar la habilidad motriz de salto en cada grupo de edad y entre los dos sexos.

* Departamento de Expresión Musical y Corporal. Facultad de Educación. C/ Rector Royo Villanova s/n. Universidad Complutense de Madrid. 28040 Madrid. E-mail: jcastejo@eucmax.sim.ucm.es

Este artículo es parte de un trabajo de investigación dirigido por este autor y financiado por el CIDE como Ayuda a la Investigación Educativa 1994 (Res. de 30-11-93, B.O.E. 14-12). En esta investigación figuran como colaboradores principales Dª Emilia Fernández García, Dª Mª Luisa Gardoqui Torralba y D. Fernando Sánchez Bañuelos.

El autor desea agradecer a D. Rafael Guzmán Murillo su aportación en la redacción de este artículo.

BEST COPY AVAILABLE

ABSTRACT

It has been designed a tests with growing difficulty level, continuing the Rasch probability model, with the objective of evaluating the motor jump ability (content from the Physical Education area) of the Primary Education students.

The statistic used for this type of tests confirm the validity and reliability of the same, so that for its application has been elected a pupils sample of the corresponding ages to the Primary Education stage (n=1951; 1057 kids and 894 girls). The results have permitted to elaborate tasks with three difficulty levels (high, mean and low) to evaluate the motor jump ability in each group of age and between the two sexes.

1. INTRODUCCIÓN

La Educación Física en la Educación Primaria ofrece una amplia variedad y opciones para lograr los objetivos generales propuestos en la LOGSE. En los contenidos y los criterios de evaluación se exponen diferentes formas y atienden, de manera más o menos acertada, a las habilidades motrices de los alumnos de estas edades. Estas habilidades expresadas en el currículum (BOE, 1992) tienen un marcado carácter cualitativo, que también se confirma desde el punto de vista pedagógico (Cratty, 1982; Gallahue, 1982; Rigal, 1987; Ruiz Pérez, 1987; Singer, 1986).

Si observamos los bloques de contenidos del área, el denominado «El cuerpo: habilidades y destrezas» contiene el salto entre sus habilidades. Del mismo modo, entre los criterios de evaluación, el nº 3 incide en: «saltar coordinadamente batiendo con una o ambas piernas en función de las características de la acción que se va a realizar». El salto será, por lo tanto, una capacidad de especial importancia a lo largo de la etapa de Educación Primaria.

Las habilidades y destrezas básicas, a diferencia de las habilidades motrices específicas, son comunes a todos los sujetos, porque desde el punto de vista filogenético, han permitido la supervivencia del ser humano, son la base de posteriores aprendizajes motores y tienen un carácter funcional (Ruiz Pérez, 1987). Con objeto de clasificarlas, estas habilidades se pueden agrupar bajo tres diferentes criterios: a) habilidades motrices basadas en la locomoción; b) habilidades motrices basadas en el manejo y dominio del cuerpo en el espacio; y c) habilidades motrices basadas en la manipulación de móviles.

Concentrando nuestra atención en el segundo apartado, el salto es una habilidad que consiste en el manejo y dominio del cuerpo en el espacio, que a su vez puede verse condicionado con la locomoción y el manejo de móviles. Sin embargo, desde un punto de vista singular, el salto se observa según el grado y control que se tiene del cuerpo en el espacio. Es por eso que en el ámbito educativo, y atendiendo al proceso madurativo, necesitamos conocer qué dominio tiene el alumnado a lo largo de la edad escolar, para presentar diferentes tareas que permitan afianzar esta habilidad con el fin de utilizarse dentro de un marco de aprendizaje significativo.

Pero no sólo debemos atender el aprendizaje, ya que el profesorado de Educación Física se encuentra con la necesidad de conocer el nivel de su alumnado para desarrollar su actividad adecuadamente (BOE, 1992; Coll, 1992; MEC, 1992). Como estas habilidades tienen un carácter procedimental, es necesario que se puedan abordar atendiendo a un orden progresivo adecuado al proceso madurativo del alumnado, con el apropiado componente de significación y funcionalidad, para que dicho alumnado pueda desarrollar su actividad de manera independiente (Zabala, 1993).

Para comprobar la adquisición de las habilidades motrices existen diferentes propuestas, pero en la mayoría de los casos, estas pruebas se han dirigido al aspecto cuantitativo del salto (distancia, número de repeticiones, etc.), y no así en cuanto al dominio de la propia habilidad (coordinación, equilibrio, etc.), como es el caso del AAHPER Test y el CAHPER Fitness Performance Test Bateries (Litwin y Fernández, 1982), y la batería Eurofit, diseñada por el Comité para el Desarrollo del Deporte en el Consejo de Europa (Council of Europe, 1981; Council of Europe, 1982a; Council of Europe, 1982b; Council of Europe, 1983), que se ha utilizado en nuestro país en diferentes estudios de investigación y aplicación (Blázquez, 1990).

Cuando se ha afrontado el componente cualitativo de esta habilidad, aunque hay propuestas interesantes que han ofrecido ayuda al profesorado, como es el caso de los trabajos de Picq y Vayer (1969), Bucher (1973) y Vayer (1977). Su aplicación no ha seguido el mismo rigor como en los casos antes mencionados, de manera que las diferentes opciones de tests y pruebas nos presentan tareas reducidas y en muchas ocasiones, no se han podido aplicar sobre la población escolar, sino en situaciones limitadas que no permiten una generalización. Por ejemplo, Litwin y Fernández (1982) muestran dos baterías de tests que sirven para medir las habilidades motrices en la enseñanza primaria, el Test de Habilidad Motora para Primaria y el Iowa Brace Test. En ambos casos, el conjunto de tareas se dirigen a todas las habilidades, comprendiendo siete y 21 pruebas respectivamente, y sólo algunas de ellas se emplean para la habilidad de salto. Blázquez (1990) también expone baterías de tests para la evaluación de la habilidad motriz en general, pero que no son bastantes para profundizar hasta un nivel que aclare el dominio que tiene el alumnado y que permita una aplicación conveniente del proceso de enseñanza aprendizaje. En cuanto a los resultados de otros estudios, no parece que sean lo suficientemente amplios como para poder generalizar (Graham, 1991; Roca, Cid e Iriarte, 1982). En suma, nos encontramos, con la necesidad de poseer instrumentos que ayuden al profesorado de Educación Física a evaluar de una manera precisa la progresión en el ámbito de las habilidades motrices, y en el caso que nos ocupa, de la habilidad de salto.

2. LAS HABILIDADES Y DESTREZAS MOTRICES BÁSICAS EN EDUCACIÓN FÍSICA: EL SALTO

El salto consiste en una batida con las piernas que hace que el cuerpo se quede suspendido en el aire por muy breves décimas de segundo, para volver a caer. Se encuentra influido por el desplazamiento, por lo que los niños de edades tempranas adquieren esta capacidad simultáneamente, aunque tienden a dominar mejor los des-

plazamientos que los saltos. Si en los desplazamientos, fundamentalmente la carrera, el equilibrio tiene importancia, en el salto, en toda su extensión, este mismo equilibrio es trascendental.

La habilidad de salto supone una actividad donde la experimentación y utilización del cuerpo en el espacio sirve para comprobar diferentes situaciones y formas de ejecución, permitiendo que el aprendizaje sobre el dominio corporal evolucione, asimilándose como una de las más importantes formas de movimiento, y aunque esta habilidad tiene un marcado componente filogenético, el aprendizaje permite que se vaya afianzando. Es decir, el salto se desarrolla y afirma mediante una correcta enseñanza, pasando de una ejecución imprecisa inicial a una soltura mayor en su etapa final. Un adecuado control de la actividad y del equilibrio permite que este salto sea considerado como óptimo, de acuerdo a diferentes interpretaciones.

Para poder introducir diferentes opciones en esta habilidad tenemos que observar las variables que la componen: impulso, vuelo y caída. Estas tres fases suponen una actuación con principio y fin, es decir que una vez se ha comenzado implica una continuidad sin interrupción. Según Wickstrom (1990), el salto puede clasificarse según el tipo y la dirección. Según el tipo podrá realizarse con una o dos piernas y la dirección puede ser adelante-atrás, arriba-abajo o lateral. Además, se puede añadir la posibilidad de realizar saltos rítmicos. Según la terminología de control y extensión de los movimientos utilizada por Singer (1986), el salto entraría dentro de las habilidades discretas, aunque los saltos sucesivos podrían considerarse como tareas continuas.

Si analizamos los tres componentes citados, en la fase de impulso es cuando se proporciona una fuerza sobre uno o dos pies para elevar al cuerpo del suelo, este impulso puede verse acompañado con la ayuda de la carrera, pero en el caso de no utilizar la carrera es necesaria la ayuda de todo el cuerpo, en especial la acción de los brazos. Los antecedentes sobre el análisis biomecánico del salto (Dyson, 1978; Hochmuth, 1973) demuestran esta ayuda de los brazos y el papel que juega el equilibrio durante el mismo. También desde el punto de vista del desarrollo, se ha observado lo esencial que es la ayuda del movimiento de los brazos en la ejecución de esta habilidad, así como la necesidad de mantener el cuerpo adecuadamente para evitar que el salto sea descoordinado (Cratty, 1982; Ruiz Pérez, 1987; Wickstrom, 1990).

En la segunda fase, durante el vuelo, el cuerpo se encuentra suspendido, permitiendo que puedan introducirse otros elementos que repercuten en la dificultad del mismo, ya que vuelve a ser necesario mantener el equilibrio pensando en la caída. En el vuelo también es conveniente un control y ayuda de los brazos para que sea uniforme.

En la fase tercera, la caída, el cuerpo vuelve a tomar contacto con la superficie sobre la que se desea caer. En esta fase es necesario el control corporal para evitar una caída desequilibrada que repercuta, bien en la acción que se ha realizado, o bien en otro movimiento que tenga que realizarse a continuación. Por regla general, la caída suele realizarse sobre los pies, no obstante, también se puede caer con otras partes del cuerpo, comprometiendo la siguiente actuación.

Desde el punto de vista madurativo y del aprendizaje, el salto puede tener diferentes dificultades, de manera que un salto para un niño de diez años puede ser muy sencillo y ese mismo salto para un niño de seis años puede ser muy complejo (Gall-

ahue, 1982). Es necesario, por lo tanto, establecer niveles de dificultad para diversificar los diferentes saltos que pueden realizarse en cada grupo de edad y la progresión que debería seguirse para emplear un nivel de dificultad lógico.

Cratty (1982) señala que a los seis años, los niños son capaces de realizar saltos rítmicos con un pie, pasando de uno a otro sin salir de su lugar. Esto permite que se pueda evaluar el salto de forma que se observe si los niños son capaces de realizar saltos con uno o ambos pies y cayendo dentro de lugares determinados. Este mismo autor presenta la evolución de la capacidad de saltar en altura y longitud, según aumenta la edad. Los aspectos madurativos y la experiencia, ya sea en el contexto escolar o no, también influyen en el dominio de esta habilidad, unidos a otros aspectos de la motricidad (Ruiz Pérez, 1987).

Pero la posibilidad de mejorar la ejecución del salto no depende sólo de aspectos evolutivos, aunque éstos deban estar presentes para poder avanzar en su mejora. Como elemento de la coordinación dinámica general, Le Boulch (1986) presenta las tareas de salto como ejercicio que implica el control de los desplazamientos en cuanto a la percepción espacio-temporal en los niños entre los seis y ocho años. Estas tareas se utilizan para superar obstáculos en diversos recorridos. Cuando los alumnos son más mayores, las tareas aumentan en dificultad, incluyendo el desplazamiento como ayuda para aumentar la capacidad de salto.

La progresión en la habilidad de saltar ha sido considerada por varios autores. Wickstrom (1990) presenta una evolución de esta habilidad, desde la edad de dos años, en la que los niños y las niñas pueden saltar desde una altura de 30 cm. con un pie adelantado, hasta la edad de tres años y medio en la que consiguen saltar a la pata coja sobre un pie de una a tres veces. Este mismo autor (Wickstrom, 1990) presenta una progresión en la dificultad de realización de salto que puede observarse en la tabla 1.

TABLA 1
TIPOS DE SALTOS QUE CONSIGUEN LOS NIÑOS EN TÉRMINOS DE DIFICULTAD PROGRESIVA SEGÚN WICKSTROM (1990, p. 83)

Salto hacia abajo con un pie, cayendo sobre el otro
Salto hacia arriba con los dos pies, cayendo sobre ambos
Salto hacia abajo con un pie, cayendo sobre ambos
Salto hacia abajo con dos pies, cayendo sobre ambos
Carrera y salto hacia delante con un pie, cayendo sobre el otro
Salto hacia delante con los dos pies, cayendo sobre ambos
Carrera y salto hacia delante con un pie, cayendo sobre ambos
Salto por encima de un objeto con dos pies, cayendo sobre ambos
Salto sobre el mismo pie de modo rítmico

Otro ejemplo sobre la evaluación del salto es el propuesto por Da Fonseca (1988), que presenta unas pruebas donde se trata de comprobar el grado de adquisición que tienen los alumnos en distintas tareas. El autor propone una serie de escalas de

identificación del desarrollo motor entre las que incluye una evolución entre los seis y los ocho años con la progresión en dificultad que puede observarse en la tabla 2 y en las que se mezclan elementos cualitativos y cuantitativos de esta habilidad. Sin embargo, creemos que hay que utilizar distintos criterios para estos elementos, ya que las diferencias de talla en estas edades puede significar que se consiga realizar o no algunas de las tareas señaladas. Sí es muy interesante su propuesta de verificar una correcta acción durante la ejecución, ya que esto permitirá realizar un perfil perceptivo motor en las edades de estudio.

TABLA 2
DIFERENTES TAREAS DE SALTO EN PROGRESIÓN DE DIFICULTAD CON NIÑOS Y NIÑAS ENTRE LOS 6 Y LOS 9 AÑOS SEGÚN DA FONSECA (1988)

6 años: se intenta identificar si salta desde 40 cm. cayendo con pies juntos.
6-7 años: salto con dos pies juntos desde 1 m.
6-7 años: salto vertical de 20-25 cm.
6-7 años: Saltos continuos con un pie al frente y el otro manteniéndole detrás, cambiando de pie cada 5 saltos.
7-8 años: saltos de obstáculos de 30-35 cm. de altura.
8-9 años: carrera a la pata coja y saltando.

La comprobación del aumento de la capacidad de salto desde el punto de vista cuantitativo progresa con la edad (Beunen, 1989; Manila y Bouchard, 1991) y lo hace de forma lineal, diferenciándose en las niñas alrededor de la pubertad. Ruiz Pérez (1987) señala que en los saltos con un apoyo y entre los seis y los nueve años, las niñas suelen aventajar a los niños. En cuanto a los saltos en longitud y altura, los niños tienden a superar a las niñas en toda la etapa primaria, haciéndose más patente con la pubertad. En cualquier caso, este componente de superioridad de los chicos sobre las chicas hay que circunscribirlo al aspecto cuantitativo del salto, es decir más alto o más lejos. No obstante, los autores siguen enfatizando que en edades anteriores a la pubertad, las diferencias apenas son significativas en relación con el sexo (Rowland, 1990), aunque en el caso del salto coordinado, las niñas superan a los niños.

Además de las diferencias entre sexos, otros factores influyen en gran medida en ésta y otras habilidades motrices. Aspectos como la madurez mental, la estatura, el peso y la composición corporal, la práctica extraescolar, la desnutrición o sobrealimentación, y diferentes factores sociales, afectan significativamente al conjunto de las habilidades motrices básicas (Manila y Bouchard, 1991).

El salto también tiene otros apartados que pueden hacerlo más complejo. En este caso dependerá de los cambios en el medio ambiente, lo que implica una adaptación de la habilidad a dichos cambios. Sin embargo, este aspecto ofrece complicaciones a la hora de estudiarlo, debido a que las variaciones medioambientales pueden ser ilimitadas y por lo tanto, también infinitas las adaptaciones que pueden realizarse.

Como síntesis de estas aportaciones, entendemos que para comprobar y evaluar la habilidad de salto, habrá que constatar el control en sus tres fases, impulso, vuelo y

caída, manifestando la armonía suficiente para que sea definido como un salto equilibrado y coordinado en toda su ejecución, sin que la postura sufra alteraciones ostensibles. Es conveniente también incluir un índice de dificultad creciente, de manera que pueda diferenciarse el dominio de la habilidad respecto a la maduración y el aprendizaje, esta dificultad está determinada por variaciones en algunas de las tres fases anteriormente mencionadas. Por ejemplo, si a un niño se le pide que realice una batida sobre los dos pies para caer sobre uno sin que pierda el equilibrio en la caída, necesitará saber que dependerá de cómo impulse y cómo se desarrolle la fase de vuelo para caer con un pie y mantener esa posición.

Por lo tanto, es necesario establecer una serie de pruebas que nos permitan evaluar de una manera adecuada, la habilidad de salto con los alumnos y alumnas de Educación Primaria, identificando las principales variables que pueden modificar y/o incrementar la dificultad en dicha habilidad, pero que a su vez nos permitan comprobar el grado de dominio para que sirva de ayuda al profesorado para comenzar o afianzar el proceso de enseñanza aprendizaje de las habilidades motrices básicas.

Para poder elaborar este conjunto de pruebas que nos permitan comprobar el nivel de adquisición de esta habilidad, hemos optado por el modelo de Rasch, por ser, bajo nuestro criterio, el que mejor objetividad nos puede proporcionar en la medida y para discriminar adecuadamente respecto al conjunto de sujetos a los que se les aplica una prueba como la aquí diseñada y aplicada (Ludlow y Haley, 1995. Özçelik y Berberoglut, 1991. Rigdon y Tsutakawa, 1987. Smith, 1994. Smith, 1996. Smith y Kramer, 1992). El modelo de Rasch es un modelo probabilístico que propone que la posibilidad de una respuesta correcta a un conjunto de ítems dicotómicos depende de la habilidad de la persona y de la dificultad del item. Mediante este modelo se ha optado por un conjunto de items que suponen una progresión de dificultad para discriminar al conjunto de personas que han participado en la aplicación del test, tal y como ha ocurrido en otro tipo de pruebas (El-Korashy, 1995. Molenaar y Hoijtink, 1996. Müller, 1987. Rost y Von Davier, 1994. Zwinderman, 1995).

Un trabajo antecedente que ha proporcionado información respecto a la utilidad del incremento de la dificultad en las habilidades motrices básicas, siguiendo el modelo de Rasch, para verificar el dominio de las mismas, es el desarrollado por Sánchez Bañuelos (1982) denominado «Aplicación de un modelo matemático probabilístico a la evaluación de la habilidad motriz en el manejo de móviles». Este trabajo, aunque dirigido a la clasificación de tareas con móviles, permite emplear su tratamiento a otras habilidades motrices. Parte de la concepción de dicho trabajo va a ser empleada en el presente documento dirigido a la habilidad de salto.

3. OBJETIVOS

Se pretende construir un instrumento que permita la evaluación del grado de adquisición y dominio de la habilidad motriz básica de salto para el alumnado de Educación Primaria, de forma que pueda conocerse, para cada edad (y si fuera necesario, el sexo), los niveles de dificultad que deben ser abordados por el profesorado para establecer la práctica durante el proceso de enseñanza aprendizaje. Cada ejercicio

utilizado deberá considerarse excluyente para permitir que con la escala se determine el lugar que ocupan los alumnos y alumnas que llegan a realizar un determinado número de ejercicios de la misma, de forma que el estadístico expuesto va a ser el del porcentaje de sujetos que han demostrado ser capaces de ejecutar correctamente las diferentes tareas que se propongan en dicha escala. Este planteamiento se basa en los objetivos y criterios de evaluación en el área de Educación Física en la Educación Primaria propuesto en la LOGSE. Resumiendo, se pretenderá:

1. Identificar las tareas relacionadas con la habilidad de salto a incluir en la escala con nivel de dificultad creciente.
2. Diseñar, validar, estandarizar y aplicar la escala para la evaluación de la habilidad motriz señalada en una población representativa.

4. METODOLOGÍA DE LA INVESTIGACIÓN

Se diseña una lista de tareas con un nivel creciente de dificultad que representa la probabilidad de que un alumno sea capaz o no de realizar un determinado ejercicio vinculado con la habilidad de salto. Esta lista de tareas tiene un carácter dicotómico (realización adecuada-realización inadecuada), lo cual supone, en el caso de cada tarea, dos únicas alternativas de valoración. Este aspecto, como se ha comentado, sigue el modelo probabilístico de Rasch para la construcción de pruebas (Jansen, 1995. Meiser y col., 1995. Zwick, Thayer y Wingersky, 1995), y que tiene aplicación a otras áreas, como puede ser el razonamiento (Molenaar y Hoijtink, 1996), actitudes (Ludlow y Bell, 1996), rendimiento (Engelhard, 1996), independientemente de que puedan existir puntualizaciones a la utilidad y aplicabilidad del modelo (Dodd y Koch, 1994. Weitzman, 1996).

El proceso de validación de las tareas propuestas, como se comprobará en este estudio, consta de los siguientes pasos:

a) Validez. Justificación de la validez intrínseca de las listas de tareas propuestas, comprobando la diferencia de dificultad entre tareas.
b) Análisis de la fiabilidad. Comprobación de que el conjunto de tareas propuestas en la lista configuran una dimensión bien definida.

Una primera lista de tareas se aplicó en un estudio piloto reducido, en un único centro elegido al azar, donde se pusieron en práctica y se analizaron los resultados para distinguir aquellas tareas que no discriminaban. El listado de tareas definitivo se pone en práctica con una muestra representativa de la población de Educación Primaria de la Comunidad de Madrid, eligiendo los individuos al azar para poder obtener una representación válida de la citada población escolar.

5. MUESTRA

La población a la que se dirige esta investigación son los alumnos y alumnas de Educación Primaria. Como muestra se elige la población escolar de la Comunidad de Madrid y se pretende mantener un margen de error del 2,5% y un intervalo de confianza del 95,5%. Esta muestra se encontrará por encima de los 1.600 alumnos,

entre niños y niñas. En total han participado 1.951 sujetos de edades comprendidas entre los 6 y los 12 años, de 45 centros de Educación Primaria. Dentro del grupo, 1.057 (54,2% de la población) son chicos y 894 (45,8% de la población) son chicas.

6. PROCEDIMIENTO

Después del análisis de la habilidad de salto atendiendo a los aspectos anteriormente señalados, se plantearon los ejercicios de forma que se estableciera, en teoría, un incremento en la dificultad de cada actividad propuesta. Para ello se procedió al diseño de una escala provisional determinando un total de 18 ejercicios como prueba piloto. El objetivo de este estudio piloto era afirmar o corregir aspectos claves de la investigación (agrupamiento de tareas, secuencia de tareas, etc.). Este estudio se realizó con una población de 120 alumnos (20 alumnos por curso, diez chicos y diez chicas) de un único centro, pero abarcando todas las edades de Educación Primaria, donde todos realizaban cada una de las pruebas en distintos momentos.

Posteriormente, y tras el análisis de las pruebas del estudio, se elaboró un listado definitivo de la escala de evaluación de la habilidad de salto. Esta nueva escala se aplica en los centros elegidos y los alumnos comienzan su ejecución en distintos tramos de esta escala, de forma que los alumnos de primer curso comienzan por la prueba uno, que según el estudio piloto es la más sencilla, sin embargo, los alumnos y alumnas de sexto curso comienzan por la prueba cuatro, según se desprende del citado estudio.

Los evaluadores, previamente instruidos, tienen que anotar en la hoja de registro, si los alumnos realizan correctamente o no la prueba presentada, tal como está redactada, de manera que se eviten interpretaciones sobre la ejecución.

En todo momento, y para evitar interpretaciones, la información que se proporcionaba al alumnado era la misma, de manera que no existieran errores que pudieran achacarse a la comprensión de la actividad que tenían que realizar. Los alumnos se sitúan por orden alfabético y todos ellos pasan por una misma prueba, y cuando han terminado ésta, vuelven a empezar con la siguiente. Cada alumno tiene un segundo intento en caso de errar en el primero.

7. ANÁLISIS DE LOS RESULTADOS

El listado de pruebas diseñado para el estudio piloto es el que constan en la tabla 3, fundamentado en diferentes publicaciones y otros documentos y pruebas que se utilizan habitualmente por el profesorado, con el fin de comprobar si se seguía la dirección adecuada. Estas 18 pruebas suponen un número elevado de tareas para que el profesorado pudiera poner en práctica esta escala en su propio centro escolar, pero sirven de base al estudio definitivo.

TABLA 3

TAREAS SELECCIONADAS PARA EL ESTUDIO PILOTO CON EL FIN DE COMPROBAR LA HABILIDAD DE SALTO CON LA POBLACIÓN DE EDUCACIÓN PRIMARIA

1. Realizar un salto con dos pies en vertical y caer con un pie y quedarse estático.
2. Salto con un pie en vertical superando una cuerda situada a la altura de la rodilla y caer al suelo con el contrario y quedarse estático.
3. Saltar desde el suelo un obstáculo de una altura igual a la longitud de su pierna (del suelo a la rodilla) con dos pies y caer al otro lado con dos pies y quedarse estático.
4. Saltar en longitud con un pie y caer con el mismo, cambio con un salto al otro pie y caer con el mismo un total de 4 saltos (es decir, izquierdo-izquierdo, derecho-derecho, izquierdo-izquierdo, derecho-derecho).
5. Paso y salto, cambio y paso y salto con un total de 4 saltos.
6. Salto a la comba individual con dos pies en el sitio un total de 12 saltos seguidos sin interrupción.
7. Salto vertical con dos pies y caer con uno, salto con este pie y caer con dos y quedarse estático.
8. Salto con dos pies desde lo alto (altura desde la mitad del muslo) y caer con dos pies y quedarse estático.
9. Salto con dos pies desde lo alto (altura desde la mitad del muslo) y caer al suelo con un pie y quedarse estático.
10. Salto desde el trampolín (con impulso de carrera de un máximo de 10 m.) con dos pies a superar un obstáculo con una altura de la longitud de la cadera y caer al suelo con dos pies y quedarse estático.
11. Salto con un pie desde el suelo y cae con dos a una altura (altura de la rodilla), desde esta altura salta con los dos pies y cae al suelo con el pie contrario del que realizó el primer impulso.
12. Salto con dos pies desde una altura (altura de la cadera) a caer en otro punto algo más bajo (altura de la rodilla) también con dos pies y desde aquí al suelo, quedando estático. Esta acción es seguida, sin interrupciones.
13. Salto a la comba con un pie, alternando (sin saltillo intermedio), 12 saltos seguidos sin interrupción.
14. Salto con un pie en corto y con el otro en largo, un total de 8 saltos.
15. Saltar desde el suelo con dos pies a caer en un pie en una altura (altura de la rodilla), con este pie saltar al suelo a caer con dos, volver a saltar con dos pies a otra altura igual a la anterior pero cae con el otro pie y de aquí al suelo cayendo con dos pies y quedarse estático.
16. Salto a la comba con dos pies, en cada salto la cuerda debe pasar dos veces, realizar un total de 10 pasadas de cuerda seguidas sin interrupción.
17. Salto a la comba hacia atrás con dos piernas, un total de 12 saltos seguidos.
18. Salto a la comba hacia atrás con un pie alternando (sin saltillo intermedio), 12 saltos seguidos sin interrupción.

Para verificar la fiabilidad interna de la prueba piloto se calculó el coeficiente alpha de Cronbach, obteniéndose, para el conjunto de las 18 pruebas del estudio entre los sujetos de la muestra que realizaron la escala, un 0.85 que se considera como un valor alto en este tipo de estudios (Morrow y col., 1995). Lógicamente, el estudio piloto ofrece unos resultados que demuestran que lo que podría considerarse teóricamente progresivo en cuanto a la dificultad, no era acertado, como puede evidenciarse en la tabla 4. Estos resultados confirmaron algunos de los ejercicios pero otros cambiaron, también se comprobó que algunos nos discriminaban entre los alumnos.

TABLA 4
PORCENTAJE DE ALUMNOS QUE LOGRAN REALIZAR CADA EJERCICIO EN EL ESTUDIO PILOTO (n=120)

Pruebas	1	3	5	8	2	4	7	9	6	12	10	11	15	14	17	13	18	16
Porcentaje	98	97	95	95	94	90	75	61	57	53	49	42	31	28	28	22	12	0

El estudio piloto termina por situar a cada ejercicio en un orden de dificultad acertado. En suma, los ejercicios, tal como se configuraron inicialmente cambian en el orden, de manera que algunas consideradas fáciles presentan mayor dificultad, y otras consideradas difíciles son más simples. Además se evidencia que algunas pruebas ofrecen una puntuación igual, lo que significa que una de ellas debe ser eliminada. Por lo tanto, se redujo el número de pruebas, para un estudio definitivo, a 15 (ver tabla 5) situando cada prueba en un nivel de dificultad más cercano a la realidad.

Con esta nueva lista, igual que en el caso anterior, se comprueba una fiabilidad con el coeficiente alpha de Cronbach, del 0.90, confirmando que el listado empleado tiene suficiente consistencia interna, derivado de la relación directa del mismo con la habilidad que se pretende evaluar.

Para comprobar la diferencia en el nivel de dificultad de cada tarea, se efectuó el cálculo del análisis no paramétrico para variables dicotómicas mediante el índice Q de Cochran, entre los diferentes ejercicios tal y como consta en la documentación para este tipo de pruebas (Thomas y Nelson, 1990), con el fin de comprobar si se cumplía o no la hipótesis de la diferenciación entre los mismos. Los resultados de estos análisis se observan en la tabla 6.

TABLA 5
LISTADO DE TAREAS DEFINITIVAS PARA LA EVALUACIÓN DE LA HABILIDAD DE SALTO EN EDUCACIÓN PRIMARIA

1. Saltar un obstáculo de una altura igual a la longitud de su pierna (del suelo a la rodilla) con dos pies y caer al otro lado con dos pies y quedarse estático.
2. Salto con un pie en vertical superando una cuerda que está a la altura de la rodilla y caer con el contrario y quedarse estático.
3. Saltar en longitud con un pie y caer con el mismo, cambio con un salto al otro pie y caer con el mismo un total de 4 saltos (es decir, izquierdo-izquierdo, derecho-derecho, izquierdo-izquierdo, derecho-derecho).
4. Salto vertical con dos pies y caer con uno, salto con este pie y caer con dos y quedarse estático.
5. Saltar con un pie un obstáculo de la altura igual a la longitud de su pierna (del suelo a la rodilla) y caer con ese mismo pie y quedarse estático.
6. Salto con dos pies desde lo alto (altura desde la mitad del muslo) y caer con un pie y quedarse estático.
7. Salto a la comba individual con dos piernas en el sitio un total de 12 saltos seguidos sin interrupción.
8. Salto con dos pies desde una altura (altura de la cadera) a caer en otro punto algo más bajo (altura de la rodilla) también con dos pies y desde aquí al suelo, quedando estático. Esta acción es seguida, sin interrupciones.
9. Salto desde el trampolín (con impulso de carrera de un máximo de 10 m.) con dos pies a superar un obstáculo con una altura de la longitud de la cadera y caer con dos pies quedando estático.
10. Salto con un pie y cae con dos a una altura (altura de la rodilla), desde esta altura salta con los dos pies y cae con el pie contrario del que realizó el primer impulso.
11. Salto desde el trampolín (con impulso de carrera de un máximo de 10 m.) con dos pies a superar un obstáculo con una altura de la longitud de la cadera y caer con un pie quedando estático.
12. Salto con dos pies a caer con un pie en una altura (altura de la rodilla), con este pie saltar al suelo a caer con dos pies, volver a saltar con dos pies a otra altura igual a la anterior pero cayendo con un pie distinto al anterior y de aquí al salto al suelo cayendo con dos pies quedando estático.
13. Salto a la comba con un pie, alternando (sin saltillo intermedio), 12 saltos seguidos sin interrupción.
14. Salto con dos pies a caer con dos pies en una altura (altura de la rodilla) y volver a caer de espaldas para caer con un pie y quedarse estático.
15. Salto a la comba con un pie, alternando hacia atrás (sin saltillo intermedio), 12 saltos seguidos sin interrupción.

TABLA 6
ÍNDICE DE COCHRAN EN LA ESCALA PARA LA EVALUACIÓN DEL SALTO DE LAS ACTIVIDADES PROPUESTAS

Sujetos	Q de Cochran	G.L.	Nivel de Sign.
1.951	11011.91	14	.000

Como procedimiento «post hoc» para determinar las diferencias entre grupos se utilizó el test de Tukey con un nivel de significación de .05 (mediante el programa estadístico SPSS 6.0 para Windows). Los resultados de dichos análisis se presentan a continuación en la tabla 7, encontrándose diferencias significativas entre los niveles escolares y la edad. Según los datos referidos, se toman como grupos de edad las siguientes: 6 años, 7 años, 8 años, 9 años, 10 años, y 11 y 12 años.

TABLA 7
ANÁLISIS DEL TEST DE TUKEY COMO PRUEBA POST HOC PARA COMPROBAR LAS DIFERENCIAS ENTRE LAS MEDIAS EN LOS GRUPOS DE EDAD

Media	NIVEL	Grp5	Grp6	Grp7	Grp8	Grp9	Grp10	Grp11	Grp12
1,00	Grp 5								
1,17	Grp 6	*							
2,20	Grp 7	*	*						
3,10	Grp 8	*	*	*					
4,18	Grp 9	*	*	*	*				
5,13	Grp10	*	*	*	*	*			
5,94	Grp11	*	*	*	*	*	*		
6,00	Grp12	*	*	*	*	*	*		

Con el objeto de comprobar si se registran diferencias significativas respecto a la habilidad motriz según el sexo, se calcularon los correspondientes análisis de diferencias entre las medias de las puntuaciones globales (número total de tareas realizadas), entre chicos y chicas. Para comprobar la hipótesis se establece el nivel de significación del 0.01. Los resultados obtenidos se presentan en la tabla 8, donde se comprueban diferencias significativas a favor de las chicas.

TABLA 8
SUMARIO DEL ANÁLISIS DE LAS DIFERENCIAS DE MEDIAS ENTRE LA PUNTUACIÓN GLOBAL SEGÚN EL SEXO

Variable	N	MEDIA	D.E.	Valor de «t»	G.L.	P
Chicos	1.057	5,80	3,80	-5,82	1936	0,000
Chicas	894	6,88	4,36			

De acuerdo con el análisis estadístico y como criterios de carácter general para realizar una adecuada utilización de las tareas de dificultad creciente con las diferentes edades y el sexo del alumnado, y en función del objetivo propuesto, se establecen los siguientes puntos:

a) Desestimar los ejercicios que son demasiado fáciles, aquellos que son realizados por más del 95% de los sujetos en cada grupo de edad; o demasiado difíciles, aquellos que no son realizados por más del 5% de sujetos de cada grupo de edad.
b) Estructurar el resto de los ejercicios en tres intervalos iguales denominados dificultad baja, cuando son realizados entre el 95% y el 65% de los sujetos en cada grupo de edad; dificultad media cuando son realizados entre el 64% y el 35% de los sujetos en cada grupo de edad; y dificultad alta cuando son realizados entre el 34% y el 5% de los sujetos en cada grupo de edad.

En la siguiente tabla, la número 9, se expone el resumen final de los ejercicios que llegan a realizar el alumnado con relación a su edad y sexo, el porcentaje de alumnos en cada edad, y su distinción entre dificultad baja, dificultad media y dificultad alta. Esta es la propuesta de los ejercicios que podrá utilizar el profesorado para evaluar la capacidad de salto en cada curso de la Educación Primaria.

Respecto a la muestra de sujetos seleccionada, y con relación a los datos obtenidos, se comprueba que hay cierta uniformidad en las tres edades iniciales respecto al primer ejercicio, el de comienzo, pero no así en la progresión de los ejercicios finales, donde se puede ver un escalonamiento, llegando las chicas en el curso tercero casi a la última prueba. A medida que aumenta la edad, el ejercicio del comienzo también se va desplazando y en quinto y sexto encontramos un punto de partida común, llegando al final en mejor situación las chicas.

En cualquier caso, y para que este listado de ejercicios tenga utilidad, existen ejercicios en los tres tramos para todas las edades y sexos, aunque es en las edades tempranas es donde encontramos que el número de ejercicios para utilizar es menor que en las edades centrales o mayores. En el curso primero encontramos un único ejercicio de dificultad baja, mientras que en dificultad media hay tres para las chicas y uno para los chicos, invirtiéndose la situación en la dificultad alta. En segundo curso encontramos la situación más igualada, de manera que existe el mismo número de

TABLA 9

DISTRIBUCIÓN DEL PORCENTAJE DE ALUMNOS Y ALUMNAS EN LOS DIFERENTES CURSOS QUE REALIZAN LAS PRUEBAS SELECCIONADAS. DA, DIFICULTAD ALTA; DM, DIFICULTAD MEDIA; DB, DIFICULTAD BAJA (ENTRE PARÉNTESIS EL PORCENTAJE DE ALUMNOS QUE REALIZAN DICHA PRUEBA)

CURSO	ALUMNOS	EJERCICIOS														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1º	CHICOS	DB (83)	DM (51)	DA (34)	DA (34)	DA (11)			DA (6)							
	CHICAS	DB (76)	DM (43)	DM (46)	DM (41)	DA (7)										
2º	CHICOS	DB (86)	DM (64)	DM (46)	DM (56)	DA (23)	DA (11)	DA (5)	DA (16)	DA (6)	DA (5)					
	CHICAS	DB (94)	DM (62)	DM (64)	DM (62)	DA (26)	DA (16)	DA (14)	DA (26)	DA (11)	DA (6)					
3º	CHICOS	DB (90)	DB (71)	DB (76)	DB (71)	DM (40)	DA (21)	DA (11)	DM (36)	DA (16)	DA (10)		DA (7)			
	CHICAS	DB (95)	DB (68)	DB (78)	DB (77)	DM (41)	DA (29)	DM (48)	DM (58)	DA (16)	DA (25)	DA (7)	DA (16)	DA (10)	DA (14)	
4º	CHICOS		DB (84)	DB (88)	DB (87)	DB (66)	DM (44)	DA (24)	DB (69)	DM (35)	DM (48)	DA (22)	DM (36)		DM (35)	
	CHICAS		DB (88)	DB (91)	DB (88)	DB (71)	DM (43)	DM (56)	DB (69)	DM (36)	DM (43)	DA (17)	DA (32)	DA (23)	DA (34)	DA (17)
5º	CHICOS					DB (85)	DM (57)	DM (39)	DB (77)	DA (34)	DM (59)	DA (24)	DM (53)	DA (10)	DM (47)	
	CHICAS					DB (80)	DM (60)	DB (75)	DB (88)	DM (59)	DM (63)	DA (19)	DM (63)	DM (45)	DM (61)	DA (19)
6º	CHICOS					DB (86)	DB (66)	DM (52)	DB (78)	DM (42)	DM (63)	DA (26)	DM (53)	DA (17)	DM (53)	DA (6)
	CHICAS					DB (88)	DB (77)	DB (82)	DB (92)	DM (52)	DB (68)	DM (38)	DM (64)	DM (39)	DM (62)	DA (14)

TABLA 10

DISTRIBUCIÓN Y EVOLUCIÓN DEL PORCENTAJE DE RESULTADOS EN LAS PRUEBAS POR LOS CHICOS ENTRE LOS SEIS CURSOS. DA, DIFICULTAD ALTA; DM, DIFICULTAD MEDIA; DB, DIFICULTAD BAJA (ENTRE PARÉNTESIS EL PORCENTAJE DE ALUMNOS QUE REALIZAN DICHA PRUEBA)

CURSO	EJERCICIOS														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1º	DB (83)	DM (51)	DA (34)	DA (34)	DA (11)			DA (6)							
2º	DB (86)	DM (64)	DM (46)	DM (56)	DA (23)	DA (11)	DA (5)	DA (16)	DA (6)	DA (5)					
3º	DB (90)	DB (71)	DB (76)	DB (71)	DM (40)	DA (21)	DA (11)	DM (36)	DA (16)	DA (10)		DA (7)			
4º		DB (84)	DB (88)	DB (87)	DB (66)	DM (44)	DA (24)	DB (69)	DM (35)	DM (48)	DA (22)	DM (36)		DM (35)	
5º					DB (85)	DM (57)	DM (39)	DB (77)	DA (34)	DM (59)	DA (24)	DM (53)	DA (10)	DM (47)	
6º					DB (86)	DB (66)	DM (52)	DB (78)	DM (42)	DM (63)	DA (26)	DM (53)	DA (17)	DM (53)	DA (6)

TABLA 11

DISTRIBUCIÓN Y EVOLUCIÓN DEL PORCENTAJE DE RESULTADOS EN LAS PRUEBAS POR LAS CHICAS ENTRE LOS SEIS CURSOS. DA, DIFICULTAD ALTA; DM, DIFICULTAD MEDIA; DB, DIFICULTAD BAJA (ENTRE PARÉNTESIS EL PORCENTAJE DE ALUMNOS QUE REALIZAN DICHA PRUEBA)

CURSO	EJERCICIOS														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1°	DB (76)	DM (43)	DM (46)	DM (41)	DA (7)										
2°	DB (94)	DM (62)	DM (64)	DM (62)	DA (26)	DA (16)	DA (14)	DA (26)	DA (11)	DA (6)					
3°	DB (95)	DB (68)	DB (78)	DB (77)	DM (41)	DA (29)	DM (48)	DM (58)	DA (16)	DA (25)	DA (7)	DA (16)	DA (10)	DA (14)	
4°		DB (88)	DB (91)	DB (88)	DB (71)	DM (43)	DM (56)	DB (69)	DM (36)	DM (43)	DA (17)	DA (32)	DA (23)	DA (34)	DA (17)
5°					DB (80)	DM (60)	DB (75)	DB (88)	DM (59)	DM (63)	DA (19)	DM (63)	DM (45)	DM (61)	DA (19)
6°					DB (88)	DB (77)	DB (82)	DB (92)	DM (52)	DB (68)	DM (38)	DM (64)	DM (39)	DM (62)	DA (14)

ejercicios para los chicos y las chicas. A partir del tercer curso, el número de ejercicios disponibles aumenta, lo que permite tener más posibilidades de elección.

Otro dato a tener en cuenta y ya mencionado, es que en la mayoría de los casos, las chicas superan a los chicos en la habilidad, y sólo cuando los chicos superan a las chicas, el porcentaje puede considerarse irrelevante. Por ello se ha separado en dos tablas, las número 10 y 11, los ejercicios para chicos y chicas.

Desde el punto de vista de la distribución de dificultad de los ejercicios de la tabla 10 correspondiente a los chicos, se puede comprobar como el ejercicio 8 cambiaría el orden hasta colocarse el sexto, el ejercicio 7 pasa a ser el 9, el 11 pasa a ser el 12 y el 13 pasa a ser el 14. En cuanto a los resultados conseguidos por los alumnos, se mantiene cierta uniformidad, de manera que las pruebas que comienzan con dificultad alta en el primer año, van perdiendo dicha dificultad cuando avanza la edad, por ejemplo la prueba 8 es dificultad alta en 1º y 2º, dificultad media en 3º, y a partir de 4º es dificultad baja. Otras pruebas se mantienen con dificultad alta durante toda la escolaridad, como la prueba 11 a partir de 4º, mientras que la prueba 15 es dificultad alta solamente en 6º, pero no aparece en ningún otro curso.

En la tabla 11, correspondiente a las chicas, y también desde el punto de vista de la dificultad inicial, algunos ejercicios cambian su posición, de forma que el 8 pasa a ser el ejercicio 6, el ejercicio 11 pasa a ser el 14 y los ejercicios 13 y 14 cambian el orden. En cuanto a los resultados conseguidos por las alumnas, observamos variaciones que hacen disminuir su estabilidad. Por ejemplo, la prueba 9 aun dentro del tramo de dificultad media, resulta más difícil para las alumnas de 6º que para las alumnas de 5º,

TABLA 12

NÚMERO DE EJERCICIOS POR EDAD Y SEXO SEGÚN EL ÍNDICE DE DIFICULTAD. DB= DIFICULTAD BAJA; DM= DIFICULTAD MEDIA; DA= DIFICULTAD ALTA. ENTRE PARÉNTESIS SE SITÚA EL EJERCICIO SEGÚN EL LISTADO DE LA TABLA 5

CURSO	SEXO	DB	DM	DA
1º	CHICOS	1 (1)	1 (2)	4 (3, 4, 5 y 8)
	CHICAS	1 (1)	3 (2, 3 y 4)	1 (5)
2º	CHICOS	1 (1)	3 (2, 3 y 4)	6 (5, 6, 7, 8, 9 y 10)
	CHICAS	1 (1)	3 (2, 3 y 4)	6 (5, 6, 7, 8, 9 y 10)
3º	CHICOS	4 (1, 2, 3 y 4)	2 (5 y 8)	5 (6, 7, 9, 10 y 12)
	CHICAS	4 (1, 2, 3 y 4)	3 (5, 7 y 8)	7 (6, 9, 10, 11, 12, 13 y 14)
4º	CHICOS	5 (2, 3, 4 5 y 8)	5 (6, 9, 10, 12 y 14)	2 (7 y 11)
	CHICAS	5 (2, 3, 4, 5 y 8)	4 (6, 7, 9 y 10)	5 (11, 12, 13, 14 y 15)
5º	CHICOS	2 (5 y 8)	5 (6, 7, 10, 12 y 14)	3 (9, 11 y 13)
	CHICAS	3 (5, 7 y 8)	6 (6, 9, 10, 12, 13 y 14)	2 (11 y 15)
6º	CHICOS	3 (5, 6 y 8)	5 (7, 9, 10, 12 y 14)	3 (11, 13 y 15)
	CHICAS	5 (5, 6, 7, 8 y 10)	5 (9, 11, 12, 13 y 14)	1 (15)

igual que la prueba 13. En la última prueba, aunque con mejores resultados que para los chicos, se constata que existe una inversión entre el curso 6º y los dos anteriores, ya que el porcentaje es menor para el curso superior. El resto de los ejercicios evolucionan normalmente, de manera que comienzan con dificultad alta en el curso 1º y terminan con una dificultad inferior en los cursos mayores.

Como resumen definitivo, exponemos en la tabla 12 los ejercicios seleccionados para evaluar el salto en cada curso según el nivel de dificultad. Lógicamente, en las situaciones en las que exista más de un ejercicio por tramo de dificultad, será labor del profesor elegir aquel que crea más oportuno, ya que no se trata de realizar todas las pruebas, sino emplear una por tramo, de manera que informe sobre qué punto se encuentra el conjunto de alumnos y alumnas que tiene a su cargo docente.

8. CONCLUSIONES

Los resultados de este estudio nos permiten corroborar, que el profesorado de Educación Física puede emplear esta escala para evaluar la habilidad de salto con sus alumnos y alumnas de Educación Primaria con el fin de diseñar las actividades que deberá desarrollar durante sus clases. La escala se ha adaptado a tres diferentes tramos de dificultad, pero que pueden discriminar sobremanera entre el grupo de alumnos según la edad y el sexo.

La propuesta original en cuanto a la dificultad teórica progresiva ha sufrido algunos cambios, provocados por los datos que ofrece la realidad con el alumnado. Sin embargo, se ofertan suficientes ejercicios dentro de los tres tramos de dificultad para que el profesorado pueda elegir aquel que considere más oportuno, aunque en algunos cursos sólo exista un único ejercicio dentro de un tramo. También es cierto que pueden encontrarse variaciones en la práctica, pero a medida que se vayan completando otros estudios se podrá generalizar en mayor proporción.

Los estadísticos aplicados confirman que la batería de ejercicios propuesta ofrecen validez y fiabilidad suficiente para poder ser aplicada en la población de estudio, representativa de la población de Educación Primaria, cuya edad se encuentra entre los 6 y los 12 años. El análisis de los datos ha demostrado también que existen diferencias entre los sexos en la habilidad de salto a favor de las chicas ya que consiguen mejores resultados en todas las pruebas y alcanzan mejor puntuación que los chicos, incluso en algunos casos llegan a rebasarlos por dos cursos de diferencia. Esto puede deberse a la influencia social y cultural o a la maduración.

BIBLIOGRAFÍA

BEUNEN, G. (1989). Biological age in pediatric exercise research. En ODED BAR-OR (ed.) *Advances in pediatric sport sciences*. (pp. 1-39). Vol. 3. Champaign, IL: Human Kinetics.

BLÁZQUEZ, D. (1990). *Evaluar en Educación Física*. Barcelona: Inde.

BOLETÍN OFICIAL DEL ESTADO (1992). nº 152.
BUCHER, H. (1973). *Trastornos psicomotrices en el niño*. Barcelona: Toray-masson.
COLL, C. (1992). *Psicología y currículum*. Barcelona: Paidós.
COUNCIL OF EUROPE (1981). *2nd European Seminar on Testing Physical Fitness*. Strasbourg: Committee for the Development of Sport.
COUNCIL OF EUROPE (1982a). *4th European Seminar on Testing Physical Fitness*. Strasbourg: Committee for the Development of Sport.
COUNCIL OF EUROPE (1982b). *Evaluation of Motor Fitness*. Strasbourg: Committee for the Development of Sport.
COUNCIL OF EUROPE (1983). *Testing Physical Fitness. EUROFIT. Experimental Battery*. Strasbourg: Committee for the Development of Sport. (provisional handbook).
CRATTY, B. J. (1982). *Desarrollo perceptual y motor en los niños*. Barcelona: Paidós.
ENGELHARD, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Education Measurement*, 33 (1), 56-70.
DA FONSECA, V. (1988). *Ontogénesis de la motricidad*. Barcelona: Núñez.
DODD, B.G. y KOCH, W.R. (1994). Item and scale information functions for the successive intervals Rasch model. *Educational and Psychological Measurement*, 54 (4), 873-885.
DYSON, G. (1978). *Mecánica del atletismo*. Madrid: INEF.
EL-KORASHY, A.F. (1995). Applying the Rasch model to the selection of items for a mental ability test. *Educational and Psychological Measurement*, 55 (5), 753-763.
GALLAHUE, D. (1982). *Understanding motor development in children*. New York: John Wiley and sons.
GRAHAM, G. (1991). Results of motor skill testing. *Journal of Teaching in Physical Education*, 10 (4), 353-374.
HOCHMUTH, G. (1973). *Biomecánica de los movimientos deportivos*. Madrid: INEF.
JANSEN, M.G.H. (1995). The Rasch poisson counts model for incomplete data: An application of the EM algorithm. *Applied Psychological Measurement*, 19 (3), 291-302.
LE BOULCH, J. (1986). *La educación por el movimiento en la edad escolar*. Barcelona: Paidós.
LITWIN, J. y FERNÁNDEZ, G. (1982). *Evaluación y estadísticas aplicadas a la Educación Física y el deporte*. Buenos Aires: Stadium.
LUDLOW, L.H. y BELL, K.N. (1996). Psychometric characteristics of the attitudes toward Mathematics and its teaching (ATMAT) scale. *Educational and Psychological Measurement*, 56 (5), 864-880.
LUDLOW, L.H. y HALEY, S.M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement*, 55 (6), 967-975.
M.E.C. (1992). *Primaria. Educación Física* (cajas rojas). Madrid: Autor.
MANILA, R.M. y BOUCHARD, C. (1991). *Growth, maduration and physical activity*. Champaign, IL: Human Kinetics.
MEISER, T., HEIN-EGGERS, M., ROMPE, P. y RUDINGER, G. (1995). Analyzing homogeneity and heterogeneity of change using Rasch and latent class models: A comparative and integrative approach. *Applied Psychological Measurement*, 19 (4), 377-391.

MOLENAAR, I.W. y HOIJTINK, H. (1996). Person-fit and the Rasch model, with an application to knowledge or logical quantors. *Applied Measurement in Education*, 9 (1), 27-45.

MORROW, J.R., JACKSON, A.W., DISCH, J.G. y MOOD, D.P. (1995). *Measurement and evaluation in Human Performance*. Champaign, IL: Human Kinetics.

MÜLLER, H. (1987). A Rasch model for continuous ratings. *Psychometrika*, 52 (2), 165-181.

ÖZÇELIK, D.A. y BERBEROGLU, G. (1991). Contributions of the Rasch model to objectivity in measurement. *Studies in Educational Evaluation*, 17, 167-188.

PICQ, L. y VAYER, P. (1969). *Educación psicomotriz y retraso mental*. Barcelona: Científico-médica.

RIGAL, R. (1987). *Motricidad humana. Fundamentos y aplicaciones pedagógicas*. Madrid: Pila Teleña.

RIGDON, S.E. y TSUTAKAWA, R.K. (1987). Estimation for the Rasch model when both ability and difficulty parameters are random. *Journal of Educational Statistics*, 12 (1), 76-86.

ROCA, J., CID, C. e IRIARTE, I. (1982). Registres sobre evolució motora (una mostra). *Apunts d'Educació física i medicina esportiva*, 19 (3), 19-38.

ROWLAND, T.W. (1990). *Exercise and children's health*. Champaign, IL: Human Kinetics.

ROST, J. y VON DAVIER, M. (1994). A conditional intem-fit index for Rasch model. *Applied Psychological Measurement*, 2 (18), 171-182.

RUIZ PÉREZ, L.M. (1987). *Desarrollo motor y actividades físicas*. Madrid: Gymnos.

SÁNCHEZ BAÑUELOS, F. (1982). Aplicación de un modelo matemático probabilístico a la evaluación de la habilidad motriz en el manejo de móviles. Memoria de Licenciatura no publicada.

SINGER, R.N. (1986). *El aprendizaje de las acciones motrices en el deporte*. Barcelona: Hispano Europea.

SMITH, R.M. (1994). A comparison of the power of Rasch total and between-item fit statiscs to detect measurement disturbances. *Educational and Psychological Measurement*, 54 (1), 42-55.

SMITH, R.M. (1996). A comparison of the Rasch separate calibration and between-fit methods of detecting item bia. *Educational and Psychological Measurement*, 56 (3), 403-418.

SMITH, R.M. y KRAMER, G.A. (1992). A comparison of two methods of test equating in the Rasch model. *Educational and Psychological Measurement*, 52, 835-846.

THOMAS, J.R. y NELSON, J.K. (1990). *Research methods in Physical Activity*. Champaign, IL: Human Kinetics.

VAYER, P. (1977). *El niño frente al mundo*. Barcelona: Científico-médica.

WEITZMAN. R.A. (1996). The Rasch model plus guessing. *Educational and Psychological Measurement*, 56 (5), 779-790.

WICKSTROM, R. (1970). *Fundamental motor patterns*. Filadelfia: Lea and Febiger. (hay traducción al castellano (1990). Patrones motores básicos. Madrid: Alianza Deporte).

ZABALA, A. (1993). *Cómo trabajar los contenidos procedimentales en el aula.* Barcelona: Grao.

ZWICK, R., THAYER, D.T. y WINGERSKY, M. (1995). Effects of Rasch calibration on ability and DIF estimation in computer-adaptive tests. *Journal of Education Measurement*, 32 (4), 341-363.

ZWINDERMAN, A.H. (1995). Pairwise parameter estimation in Rasch models. *Applied Psychological Measurement*, 4 (19), 369-375.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 69-82

EFECTOS DE LA DIFERENCIA DE EDAD Y SEXO ENTRE TUTOR Y TUTORADO EN UNA INTERVENCIÓN COGNITIVO-CONDUCTUAL DIRIGIDA A MEJORAR LA AUTOESTIMA, BASADA EN UNA ESTRUCTURA DE TUTORÍA RECÍPROCA

*Miguel Moreno Moreno**
Campus Universitario de Cartuja. Granada

RESUMEN

Los constructos como el autoconcepto o la autoestima, con una clara dependencia del entorno humano en lo que se refiere a su formación, desarrollo y cambio, están fuertemente modulados, tanto en su formación como en su mantenimiento, por la cualidad de interacción del sujeto con los que le rodean, dependiente a su vez de las características perceptuales del individuo. Por ello, los procesos de modificación de estas variables deben verse afectadas por la simetría —asimetría en la edad y por la similitud— diferencia de sexo entre el agente social implicado en el cambio y el individuo. Para estudiar la relación entre estas dos dimensiones y la modificación de la autoestima, el autoconcepto y el autocomportamiento en un contexto interactivo, se realiza una intervención cognitivo-conductual sobre un grupo experimental, compuesto por 39 estudiantes universitarios, frente a un grupo control que no recibió tratamiento, formado por 34 individuos.

Palabras clave: *Tutoría recíproca. Autoestima. Autoconcepto. Diferencia de edad tutor-tutorado. Diferencia sexo tutor-tutorado.*

---

* Dpto. Psicología Evolutiva y de la Educación. Facultad de Ciencias de la Educación. Campus Universitario de Cartuja. 18071 GRANADA. TF: 958 24 39 67. FAX: 958 24 89 75.

ABSTRACT

Constructs such as self-concept or self-esteem, with a clear dependence of the human environment with respect to formation, development and change, ar strongly modulated, both in its formation and maintenance, by the quality of the interaction of the person with those who surround him, which itself depends on the perceptual characteristics of the person. Beacause of this, the processes of modification of these variables will be affected by simetry-asymetry in age and by the symilarity-difference in gender between the socil agent insolved in the change and the person. To study the relationship between these two dimensions and the modification of self-esteem, self-concept, and behabior in an interactive context a cognitive-behabioral intervention was conducted with an experimental group formed by 39 university students and whas compared with a control group (34 students) that did not received any treatment.

1. INTRODUCCIÓN

Las técnicas de intervención fundamentadas en modelos cognitivo-conductuales están demostrando su efectividad en la resolución de problemas diversos, directamente relacionados con la capacidad adaptativa del educando, y con aspectos tan relevantes como el rendimiento académico o la calidad de las interacciones entre el alumno y su medio. La intervención dirigida a la mejora de conductas depresivas (Dadds et al., 1997), agresivas (Phillips, Schwean y Saklofske, 1997), hiperactivas (Kaduson, y Finnerty, 1995), o trastornos de la alimentación (Eldredge et al., 1997) son sólo algunos ejemplos. Otros estudios se centran en la mejora de la autoestima en población escolar empleando terapia racional emotiva (Sapp, 1986), cognitivo-conductual (Sapp, 1994), combinación de terapia conductual y cognitivo-conductual (Valliant, 1993), o de terapia breve y cognitiva (Martin, 1993). La carencia de estudios rigurosos que analicen los elementos asociados a la efectividad de las intervenciones realizadas, dificulta la comprensión de los factores relacionados con el cambio cognitivo y comportamental. Este estudio analiza la influencia de variables asociadas al efecto de un proceso de intervención psicopedagógica centrado en la modificación del autoconcepto y la autoestima a través de un procedimiento de tutoría recíproca. Esta estructura de aprendizaje ha demostrado en numerosos estudios su efectividad para mejorar el rendimiento en el aprendizaje y, como efecto colateral, su repercusión positiva sobre la autoestima. Aunque casi todas las investigaciones mencionan si la relación tutor-tutorado era simétrica o asimétrica en cuanto a la edad y al sexo de los participantes en la interacción, ningún estudio analiza directamente los efectos directos de estas variables sobre la modificación de la autoestima.

La elección del método de intervención, cognitivo-conductual, viene sustentada en estudios como los citados anteriormente, que avalan la validez social, ecológica, y científica del mismo. Para elegir las técnicas tomamos como referencia los datos más recientes y rigurosos que conocemos acerca del desarrollo del autoconcepto y la autoestima. Estos muestran la significatividad de la dimensión afectiva en la dinámica evolutiva de la autoestima. El desarrollo de esta dimensión depende en gran medida

de ámbitos interdependientes como la escuela y la familia, así como de la influencia de agentes socializadores distintos, aunque complementarios, como los padres, el profesorado o el grupo de iguales. Todos ellos participan en el desarrollo cognitivo, afectivo y social del niño, y contribuyen sobremanera a la formación de esquemas de pensamiento que a través de su función heurística/interpretativa configuran la relación del sujeto con el medio físico y social.

La especial relevancia de las interacciones sociales del sujeto con personas significativas sugiere la idea de modificar esta última empleando, al igual que en su desarrollo, las interacciones con otro igual, facilitándose así la generalización de los nuevos pensamientos adquiridos y relacionados directamente con esta dimensión afectiva de la persona. Las experiencias ya realizadas en el ámbito del aprendizaje de tareas académicas utilizando experiencias innovadoras como la tutoría recíproca así lo sugieren (Fernández y Melero, 1995; Jacobs y James, 1994; Marshak, 1994, entre otros). Estudios como el de Verba (1998) señalan la importancia de la simetría-asimetría en las interacciones, mostrando también el efecto negativo que produce esta última al originar que tutor y tutorado focalicen la atención en elementos distintos, disminuyendo notablemente la atención que se presta a la satisfacción de las necesidades del tutorado. Nosotros hemos optado por formar parejas de igual nivel de desarrollo en conocimiento y habilidad.

El efecto positivo de este proceso interactivo podría estar asociado a la similitud o a la diferencia de sexo entre tutor y tutorado, así como a la mayor o menor diferencia de edad. Mientras unos estudios como el de Marshak (1994) encuentran razones que apoyen la diferencia de edad, otros como el King, Staffieri, y Adelgais, (1998) reflejan las ventajas de utilizar parejas constituidas por sujetos de la misma edad. La diversidad de muestras utilizadas, así como la diversidad de contenidos de aprendizaje sobre los que se centran estas investigaciones dificultan la obtención de conclusiones definitivas. Lo mismo ocurre en lo que respecta a la igualdad-similitud de sexo entre tutor y tutorado. Los estudios de Horn, Collier, Oxford, Bond, y Dansereau, 1998, o de Verba,1998 reflejan las ventajas de la asimetría en la variable género. Aunque los estudios citados refieren efectos positivos sobre la autoestima, ninguno de ellos analiza de forma conjunta los efectos que tienen las variables simetría-asimetría de sexo y edad en la pareja tutor-tutorado, en una intervención específicamente dirigida a mejorar la autoestima, el autoconcepto y el autocomportamiento.

Podríamos pensar, como consecuencia del análisis de los estudios que han utilizado la tutoría recíproca para mejorar el rendimiento en el aprendizaje, que estas variables van a influir sobre el autoconcepto y la autoestima., así como sobre el autocomportamiento. Las investigaciones arriba citadas nos llevan a plantear los siguientes supuestos. Planteamos una primera hipótesis que afirma que la mejora debe ser significativamente mayor cuando los dos miembros de la pareja sean del mismo sexo (hipótesis 1). La diferencia de sexo entre tutor y tutorado podría dificultar el proceso de autoconocimiento, condicionando o dirigiendo la búsqueda de aspectos negativos o positivos en una u otra dirección. De igual modo podríamos pensar que al ser generalmente las interacciones con el otro sexo las que más fácilmente suelen afectar a la autoestima, y no tanto al autoconcepto, interaccionar en este contexto de identifica-

ción y análisis de las experiencias vividas podría dificultar el proceso de búsqueda de justificaciones a los comportamientos autoobservados.

Autores como Fitzsimons-Lovett (1998) sugieren los aspectos positivos de la diferencia de edad tutor-tutorado, mostrando una mejora de la autoestima cuando los dos miembros de la pareja son de distinta edad. Nosotros esperamos que a mayor diferencia de edad entre los miembros de la pareja la modificación sea menor, dados los planteamientos de líneas anteriores, valorándose positivamente una similitud en la edad de los dos miembros de la pareja (hipótesis 2). Estas hipótesis deben verificarse para el autoconcepto y la autoestima, así como para el autocomportamiento, dimensiones distintas aunque interdependientes, que se verán positivamente incrementadas tras la intervención (hipótesis 3).

2. SUJETOS

Sesenta y cuatro sujetos eligieron participar de forma voluntaria en la investigación. Todos ellos fueron informados con anterioridad del objetivo general del estudio: implementar un proceso de intervención dirigido a la modificación de la autoestima por medio de una serie de actividades desarrolladas por parejas. Veinte mujeres y cuarenta y cuatro hombres se subdividían en grupos experimental y control, el primero compuesto por ocho varones y treinta y una mujeres, y el segundo por quince varones y diecinueve mujeres, con una edad media de 25.7 años. Una primera evaluación de las variables dependientes (autoconcepto, autoestima y autocomportamiento), reflejó puntuaciones en torno a la media de su edad en la mayoría de los sujetos. Tan sólo 12 sujetos puntuaron por debajo de la media en la variable autoestima. Las parejas tutor-tutorado fueron realizadas al azar para el grupo experimental, mientras que el grupo control no recibió ninguna intervención.

3. MÉTODO

3.1. Instrumentos

Para la evaluación de la autoestima se ha utilizado el Tennessee Self-Concept Scale (T.S.C.S.), elaborado por Fitts en 1965 (citado en Garanto, 1984). El instrumento se ha utilizado en España (Garanto, 1984, quien presenta la versión que aquí hemos utilizado) y en los Estados Unidos. Cuenta con tres escalas (autoconcepto, autoestima y autocomportamiento), subdividiéndose cada una de ellas en cinco dimensiones: ético-moral, personal, familiar, social y física. Además el test lleva incluida una escala de 10 ítems tomada del MMPI que mide autocrítica. Nosotros únicamente empleamos una puntuación global de cada una de las tres primeras escalas, y la puntuación total, suma de las tres anteriores, excluyendo la escala de autocrítica, asociada a características psicopatológicas y que nosotros no utilizamos al no considerar relevante esta variable para los objetivos propuestos.

A los sujetos se les repartió un documento que contenía detalladamente todas y cada una de las etapas a seguir, que explicamos en el apartado de procedimiento.

3.2. Procedimiento

La intervención realizada, basada en la interacción en parejas, con un marcado carácter psicoeducativo, se estructuró en cuatro partes, a realizar ambos miembros, los cuales intercambiaban sus roles tutor-tutorado a lo largo de todo el proceso, alternando estos roles en cada una de las actividades, y realizando estas los dos miembros de la pareja. La secuencia de tareas a realizar era la siguiente:

1) Conocimiento acerca de la distinción entre autoconcepto y autoestima

En esta primera fase, cada uno de los componentes de la pareja, previamente auto-formado sobre los conceptos e implicaciones de estos dos constructos sobre el bienestar, y sobre el rendimiento académico, formulaba una serie de preguntas a su compañero/a que reflejaran el conocimiento de cada uno de ellos acerca de esos temas, y que a la vez, le permitiese al tutor, conocer con cierta seguridad que su tutorado dominaba bien los términos objeto de estudio. Recordamos que al combinar los roles en este primer momento, al igual que en los siguientes, en un primer momento cada uno de ellos comprobaba si el compañero/a conocía los conceptos citados, y viceversa. En este documento se informaba de los elementos principales que debían conocerse: conceptualización correcta de los términos y modo en el que las distintas investigaciones relacionaban los términos autoconcepto, autoestima y autocomportamiento.

2) Evaluación del autoconcepto y de la autoestima

Una vez que los dos individuos conocían el sentido y significado de los términos tratados se realizaba una primera evaluación, previa a la fase de intervención, que permitiese a cada uno de ellos conocer el grado de autoestima autoinformado que presentaba cada uno de los miembros del par, tanto él/ella como su tutorado/tutor. Entre los dos conocían más a fondo la prueba, comprendían las instrucciones, resolvían las dudas entre los dos, y una vez que ambos estaban seguros de no tener dudas acerca del modo de contestar a la prueba de evaluación respondían al T.S.C.S., corrigiendo cada uno el de su tutorado. Al inicio de este proceso se instruyó a cada uno sobre el modo de corregir el test, proporcionando además junto a las hojas de respuestas las instrucciones de corrección.

3) Conocimiento «tutor-tutorado»

En una tercera fase el objetivo fundamental es que el tutor conozca más a su tutorado. Dado que es por parejas, ambos realizan el ejercicio alternando estos dos roles para cada etapa, en lugar de realizar primero toda la secuencia el tutor y luego intercambiando los papeles. Para ello se aconsejaba previamente a cada uno de ellos que uno contestase a las preguntas, y posteriormente, una vez finalizadas todas las cuestiones respondiese el otro miembro. Esta fase pretendía profundizar en las atribu-

ciones o interpretaciones que los individuos hacen de su conducta y que les hace sentirse bien o mal.

En esta etapa el intercambio de preguntas se podía ayudar de las siguientes cuestiones que se proporcionaban a todos los sujetos en el siguiente orden:

— «Escribe de 15 a 20 aspectos positivos de tu forma de ser».
— «Describe los aspectos negativos que menos te gustan de ti».
— «Explica a tu compañero/a por qué te gustan o no esos aspectos y pon ejemplos variados de situaciones concretas que te hacen sentir bien o mal explicando claramente por qué te hacen sentir así».

La respuesta a estas preguntas perseguía fundamentalmente comprender un poco mejor la naturaleza irracional de muchas de esas creencias acerca de sí mismo, compuestas muchas de ellas por una dosis de subjetivismo y falta de objetividad.

4) Refutación de las creencias acerca de sí mismo

Una vez conocidas por el tutor aquellos aspectos por los cuales el tutorado se siente mal consigo mismo, y las razones que da para ello, muchas de ellas totalmente subjetivas y con una gran dosis de absolutismo, cada una de ellas se iba cuestionando con el tutorado prestando una especial atención a la búsqueda de ejemplos reales de la vida del compañero que demostraran el carácter absolutista e irracional de esas creencias acerca de sí mismo, profundamente asociadas a sus interacciones con el medio social más cercano, en los casos en que procediera.

5) Evaluación post-intervención

De nuevo tutor y tutorado intercambian roles contestando de nuevo al T.S.C.S., analizando cada uno de ellos las posibles razones por las que puede haberse producido su mantenimiento, descenso o incremento de sus puntuaciones en las escalas medidas.

3. RESULTADOS

a) Efectos de la intervención sobre las variables autoconcepto, autoestima y autocomportamiento

Para comprobar los efectos de la intervención realizamos un análisis de covarianza sobre las variables de autoestima utilizando como covariado las puntuaciones obtenidas en la evaluación inicial previa a la intervención, resultando diferencias estadísticamente significativas entre las puntuaciones previas y posteriores a la intervención en todas las variables medidas.

TABLA 1
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCONCEPTO (TACPOS)

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariados	5175.998	1	1175.998	42.615	.003
TAC	5175.998	1	1175.998	42.615	
Efectos principales	957.048	1	1175.995	19.344	
GRUPO	957.048	1	957.048	19.344	.000
Explicado	6133.046	2	3066.523	61.979	.000
Residual	3018.063	61	49.746		
Total	9151.109	63	145.246		

TABLA 2
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOESTIMA (TAEPOS)

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	2297.940	1	2297.940	40.96	.000
TAE	2297.940	1	2297.940	40.96	.000
Efectos principales	763.748	1	763.748	13.599	.000
GRUPO	763.748	1	763.748	13.599	.000
Explicado	3061.688	2	1530.844	27.257	.000
Residual	3425.921	61	56.163		
Total	6487.609	63	102.978		

TABLA 3
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCOMPORTAMIENTO (TCO)

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	1018.352	1	1018.352	15.541	.000
TCO	1018.352	1	1018.352	15.541	.000
Efectos principales	2172.422	1	2172.422	33.153	.000
GRUPO	2172.422	1	2172.422	33.153	.000
Explicado	3190.744	2	1595.387	24.347	.000
Residual	3997.163	61	65.527		
Total	7187.938	63	114.094		

TABLA 4
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE PUNTUACIÓN TOTAL (TOT)

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	23838.723	1	23838.723	46.805	.000
TOT	23838.723	1	23838.723	46.805	.000
Efectos principales	10744.577	1	10744.577	30.110	.000
GRUPO	10744.577	1	10744.577	30.110	.000
Explicado	34583.300	2	17291.650	48.458	.000
Residual	21767.137	61	356.650		
Total	56350.437	63	894.451		

b) Efectos de la edad del tutor sobre la eficacia del procedimiento

Para comprobar esta hipótesis realizamos un segundo análisis de covarianza sobre las mismas variables pero utilizando como variable independiente la edad del tutor, resultando un efecto nulo de esta variable sobre las diferencias entre las puntuaciones pre y post-intervención (ver Tablas 5, 6, 7 y 8).

TABLA 5
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCONCEPTO (TAC) CON LA EDAD DEL TUTOR (EDTER) COMO VARIABLE INDEPENDIENTE

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	560.440	1	560.440	14.340	.002
TAC	560.440	1	560.440	14.340	.002
Efectos principales	287.246	6	47.874	1.225	.347
EDTER	287.246	6	47.874	1.225	.347
Explicado	847.685	7	121.098	3.099	.031
Residual	586.228	15	39.082		
Total	1433.913	22	65.178		

TABLA 6
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOESTIMA (TAE) CON LA EDAD DEL TUTOR (EDTER) COMO COVARIADO

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	967.242	1	967.242	13.716	.002
TOT	967.242	1	967.242	13.716	.002
Efectos principales	416.261	6	69.377	.984	.470
EDTER	416.261	6	69.377	.984	.470
Explicado	1383.503	7	197.643	2.803	.044
Residual	1057.801	15	70.520		
Total	2441.304	22	110.968		

TABLA 7
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCOMPORTAMIENTO (TCO) CON LA EDAD DEL TUTOR (EDTER) COMO COVARIADO

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	967.242	1	967.242	13.716	.002
TOT	967.242	1	967.242	13.716	.002
Efectos principales	416.261	6	69.377	.984	.470
EDTER	416.261	6	69.377	.984	.470
Explicado	1383.503	7	197.643	2.803	.044
Residual	1057.801	15	70.520		
Total	2441.304	22	110.968		

TABLA 8
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE PUNTUACIÓN TOTAL (TOT) CON LA EDAD DEL TUTOR (EDTER) COMO COVARIADO

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	237.980	1	237.980	3.767	.071
TOT	237.980	1	237.980	3.767	.071
Efectos principales	406.249	6	67.708	1.072	.421
EDTER	406.249	6	67.708	1.072	.421
Explicado	644.229	7	92.033	1.457	.255
Residual	947.510	15	63.167		
Total	1541.739	22	72.352		

c) Efectos de la simetría-asimetría en el género sobre la eficacia de la intervención

No se observan tampoco efectos diferenciales del tratamiento en función de esta variable, tal como reflejan las tablas 9, 10, 11, 12.

TABLA 9
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCONCEPTO (TAC) CON EL SEXO DEL TUTOR (ST) COMO VARIABLE INDEPENDIENTE

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	661.223	1	661.223	9.713	.004
TAE	661.223	1	661.223	9.713	.004
Efectos principales	189.770	1	189.770	2.788	.104
ST	189.770	1	189.770	2.788	.104
Explicado	850.993	2	425.496	6.250	.005
Residual	2450.751	36	68.076		
Total	3301.744	38	86.888		

TABLA 10
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOESTIMA (TAE) CON EL SEXO DEL TUTOR (ST) COMO VARIABLE INDEPENDIENTE

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	1599.089	1	1599.089	22.888	.000
TAE	1599.089	1	1599.089	22.888	.000
Efectos principales	250.814	1	250.814	3.590	.066
ST	250.814	1	250.814	3.590	.066
Explicado	1849.904	2	924.952	13.239	.000
Residual	2115.173	36	69.866		
Total	4365.077	38	114.870		

TABLA 11
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE AUTOCOMPORTAMIENTO (TCO) CON EL SEXO DEL TUTOR (ST) COMO COVARIADO

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	889.596	1	889.596	11.896	.001
TAE	889.596	1	889.596	11.896	.001
Efectos principales	206.127	1	206.127	2.756	.106
ST	206.127	1	206.127	2.756	.106
Explicado	1095.724	2	547.862	7.326	.002
Residual	2692.174	36	74.783		
Total	3787.897	38	99.682		

TABLA 12
ANÁLISIS DE COVARIANZA SOBRE LA VARIABLE PUNTUACIÓN TOTAL (TOT) CON EL SEXO DEL TUTOR (ST) COMO COVARIADO

Fuente de Variación	Suma de Cuadrados	G.L.	Media de Cuadrados	F	Signif. F.
Covariado	8551.657	1	8551.657	17.287	.000
TOT	8551.657	1	8551.657	17.287	.000
Efectos principales	1986.743	1	1986.743	4.016	.053
ST	1986.743	1	1986.743	4.016	.053
Explicado	10538.400	2	5269.200	10.652	.000
Residual	17808.369	36	494.677		
Total	28346.769	38	745.968		

4. DISCUSIÓN

Estudios recientes establecen diferencias importantes entre autoconcepto y autoestima, el primero referido a la dimensión cognitiva, y la segunda con un marcado carácter afectivo, dependiente en gran medida de las evaluaciones, positivas y negativas, que un sujeto realiza acerca de sí mismo (autoestima), de lo que conoce acerca de sí (autoconcepto), y de su comportamiento (autocomportamiento). Al estar la autoestima asociada tanto a las propias evaluaciones como a la información que nos proporcionan otros significativos, cabe esperar que características específicas de aquellos como la edad, el género, etc., puedan influir en los efectos de la intervención al estar asociadas a las relaciones que se establecen en el contexto social en el que se puede ver envuelto e identificados un determinado individuo. Los estudios citados en la introducción muestran resultados contradictorios a este respecto. En esta investigación observamos un efecto nulo de estas variables sobre los efectos de la intervención diseñada. Esto quiere decir que no ha influido en la mejora del autoconcepto, la

autoestima o el autocomportamiento la simetría o asimetría en la edad del tutor y tutorado, y que tampoco la similitud o diferencia de sexo entre ambos ha sido una variable que permita explicar los resultados obtenidos. Tan sólo el sexo del tutor parece tener una influencia relevante sobre la autoestima aunque sin llegar a alcanzar niveles estadísticamente significativos. Ciertamente parece que tiende a ser positiva la participación de un tutor del mismo sexo para modificar esta variable. La interacción social estructurada, en este caso por medio de una secuencia de actividades claramente definida, parece ser un contexto adecuado para la mejora del autoconcepto en la medida en que el diálogo con un igual permite conocer aquello que los demás ven y el sujeto no percibe; si el clima es adecuado y las actividades lo facilitan, el alumno puede profundizar en aspectos relacionados a su identidad que pueden estar interfiriendo en su capacidad para adaptarse de forma satisfactoria a las demandas del entorno familiar (padres), escolar (educadores) y social (relaciones entre el alumno y padres, profesores e iguales). Este proceso de reflexión acerca de sí mismo se complementa en el procedimiento aplicado con la oportunidad de ahondar, casi de forma inmediata en aspectos que el sujeto no conoce y que puede ampliar preguntando a su compañero. Si bien hemos visto que la diferencia de edad no parece importante, podemos pensar que cuanto mayor sea esta más difícil será la identificación con las experiencias de la otra persona, e incluso más reparos encontrará el alumno para realizar determinadas preguntas. Estas dificultades podrían verse incrementadas si existe además asimetría en el género. Otras razones podrían ayudar a comprender la efectividad de la intervención. Si nos centramos en los mecanismos responsables del cambio en las cogniciones, el conflicto cognitivo (en términos piagetianos), y sociocognitivo (desde un punto de vista vigotskiano), nos permite comprender la efectividad del diálogo tutor-tutorado como elemento desencadenante del cambio, al generar un desequilibrio en la estructura cognitiva del sujeto y motivar los procesos responsables del restablecimiento del equilibrio perdido. Igual ocurriría con las evaluaciones cognitivas responsables de una baja autoestima, al focalizar el individuo la atención sobre aspectos y justificaciones negativas ignorando alternativas o aspectos positivos de sí mismo. El diálogo y la oportunidad de confrontar, y no de discutir, puntos de vista distintos a los suyos, razones alternativas que él no ve, así como aspectos positivos que puede ver el compañero, es un elemento facilitador del cambio en las evaluaciones. Lógicamente es fundamental que este proceso de cambio en los aspectos cognitivos y afectivos vaya acompañado, para que sea efectivo a largo plazo, de algún método de mantenimiento, como podría ser un simple autorregistro, en el que tutor y tutorado, registren durante un tiempo aquellos tipos de situaciones, aspectos, o razonamientos, que ahoran perciben tras el proceso de intervención, una vez resuelto el conflicto y mejorado el aprendizaje (reflejado esto en la mejora de las puntuaciones en las variables medidas), hasta que esta práctica de analizar razones alternativas a los sucesos, y esta identificación de pensamientos que reducen poco a poco la autoestima sea algo habitual.

A modo de conclusión hemos de mencionar que la tutoría de pares, además de como recurso educativo, resulta efectiva para modificar aspectos cognitivos y afectivos a diferentes edades y niveles educativos. Empleada como técnica para modificar

la autoestima de forma directa resulta novedosa. Este estudio debería de complementarse con otras investigaciones que, realizadas sobre población con menor edad, repliquen los resultados obtenidos, y analicen más en profundidad los mecanismos responsables del cambio en las variables medidas, así como otro tipo de interacciones que puedan también favorecer el cambio en el sujeto.

BIBLIOGRAFÍA

DADDS, M.R.; SPENCE, S.H.; HOLLAND, D.E.; BARRETT, P.M. et al. (1997) Prevention and early intervention for anxiety disorders: A controlled trial, *Journal of Consulting and Clinical Psychology, 65*(4) 627-635.

ELDREDGE, K.L.; AGRAS, W.S.; ARNOW, B ; TELCH, C.F., et al. (1997) The effects of extending cognitive-behavioral therapy for binge eating disorder among initial treatment nonresponders, *International Journal of Eating Disorders, 21*(4) 347-352.

FERNÁNDEZ, P. y MELERO, Mª A. (1995) *La interacción social en contextos educativos.* Madrid: Siglo XXI.

FITZSIMONS-LOVETT, ANN (1998) *Enhancing Self-Respect: A Challenge for Teachers of Students with Emotional/Behavioral Disorders.* Reston, VA: Council for Children with Behavioral Disorders.

GARANTO, J. (1984) *Las actitudes hacia sí mismo y su medición.* Barcelona: Servicio de Publicaciones de la Universidad de Barcelona.

HORN, E.M.; COLLIER, W.G.; OXFORD, J.A.; BOND, C.F. Jr.; y DANSEREAU, D.F. (1998) Individual differences in dyadic cooperative learning. *Journal of Educational Psychology, 90*(1), 153-161.

JACOBS, G.M. y JAMES, J.E. (1994) *A Comparison of Workplace Groups with Groups in Education.* Comunicación presentada al Congreso Anual de Profesores de Inglés a hablantes de otras lenguas, Baltimore, 8-12 de Marzo.

KADUSON, H.G. y FINNERTY, K. (1995) Self-control game interventions for attention-deficit hyperactivity disorder, *International Journal of Play Therapy, 4*(2), 15-29.

KING, A.; STAFFIERI, A.; y ADELGAIS, A. (1998) Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*(1): 134-152.

MARSHAK, D. (1994) *From Teachers' Perspectives: The Social and Psychological Benefits of Multiage Elementary Classrooms.* Comunicación presentada a la Conferencia y Muestra anual «Emerging Images of Learning: World Perspectives for the New Millennium» (Chicago, 19-22 de Marzo).

MARTIN, J.I. (1993) Self psychology and cognitive treatment: An integration, *Clinical Social Work Journal, 21*(4), 385-394.

PHILLIPS, D.R.; SCHWEAN, V.L.; SAKLOFSKE, D.H. (1997) Treatment effect of a school based cognitive-behavioral program for aggressive children, *Canadian Journal of School Psychology, 13*(1), 60-67.

SAPP, M. (1996) Irrational beliefs that can lead to academic failure for African American middle school students who are academically at-risk, *Journal of Rational Emotive and Cognitive Behavior Therapy, 14*(2), 123-134.

SAPP, M. (1994) Cognitive-behavioral counseling: Applications for African-American middle school students who are academically at-risk, *Journal of Instructional Psychology, 21*(2), 161-171.

VALLIANT, P. (1993) Cognitive and behavioural therapy with adolescent males in a residential treatment centre, *Journal of Child and Youth Care, 8*(3), 41-49.

VERBA, M. (1998) Tutoring interactions between young children: How symmetry can modify asymmetrical interactions. *International Journal of Behavioral Development, 22*(1), 195-216.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 83-96

EVALUACIÓN DEL PROGRAMA DE ORIENTACIÓN PARA LA INSERCIÓN LABORAL EN LOS CENTROS PENITENCIARIOS

*Gemma Filella Guiu**
Universitat de Lleida

RESUMEN

En este artículo partimos de la base de que es necesaria la inserción laboral para que los internos en los centros penitenciarios puedan conseguir una inserción social óptima una vez se encuentren en libertad. Esta inserción requiere un proceso normalizador que en ningún caso tiene que empezar cuando el interno se encuentra en libertad o muy cerca de ella. En este sentido presentamos la intervención basada en tres niveles de orientación para favorecer que el interno pueda encontrar un trabajo una vez esté en libertad. En concreto presentamos los resultados de la evaluación del Programa de Orientación para la Inserción Laboral que correspondería al primer nivel de intervención. Utilizando el modelo CIPP de Sufflebeam hemos valorado la eficacia, la eficiencia, la efectividad y el impacto del programa a partir de una muestra de internos del Centro Penitenciario Ponent, cuyos resultados consideramos satisfactorios.

ABSTRACT

In this report we think that it is essential to find a job when prisoners are free to get an optimal rehabilitation.

* Departament de Pedagogia i Psicologia. Universitat de Lleida. Complex de la Caparrella s/n. 25198 Lleida. Em-mail: gfilella@pip.udl.es

BEST COPY AVAILABLE

We must teach them the necessary strategies and skills to obtain a labour market integration before they finish their sentence.

We introduce a three level guidance program and specially assessment to the first level. We have used the CIPP model from Stufflebeam and we have appreciated the efficacy, the efficiency and the effectiveness from this program which has been aplicated in a prison where we consider that the results have been satisfactory.

1. INTRODUCCIÓN

La intervención que se lleva a cabo en los centros penitenciarios tiene la finalidad de facilitar la posterior inserción social de las personas internas. Esta intervención orientadora se asume desde una visión globalizadora, la cual, a partir de las necesidades de los internos, tiene el objetivo de dotarles de las competencias necesarias para favorecer su inserción social.

Un aspecto de esta intervención sería la orientación para la inserción laboral, pues el hecho de desarrollar una actividad laboral es una de las más importantes instancias socializadoras del ser humano. Consideramos que sin una ocupación que les permita sobrevivir dignamente (aunque no obtenga una autorrealización personal) es inviable una inserción social mínimamente satisfactoria una vez hayan cumplido la sentencia.

Glasser, ya en 1964, apuntaba la necesidad de desarrollar un trabajo para hacer posible la adaptación a la sociedad de este tipo de población (Glasser en Morris, 1981).

Estudios a nivel macrosocial han comprobado que los índices de paro y de delincuencia aumentan y disminuyen paralelamente, aunque es necesario un intervalo de tiempo para que se influencien (Blanco, 1990; Cantor y Land, 1991; Córcoles, 1994; Farrigton y West, 1990; entre otros).

Sothill y Holme (1981) observaron que el trabajo favorecía la no-reincidencia, correlacionando las dos variables obtuvieron una correlación negativa, es decir a más tiempo trabajado menos volvían a delinquir las personas que habían estado recluidas en centros penitenciarios. Al contrario, comprobaron que correlacionaban positivamente el tiempo de prestación económica con la reincidencia.

Diversas investigaciones sobre programas de rehabilitación indican que el trabajo tiene un peso factorial de hasta un 20%, mientras que otras variables como la familia tienen menos incidencia en el proceso de reinserción social (Redondo, 1992), por lo tanto podemos concluir que la capacidad de desarrollar y mantener un trabajo de manera estable es uno de los factores más importantes de reinserción social.

La mayoría de autores comparten la idea de que la inserción laboral no sólo se refiere a la entrada en la vida activa sino, también, a la definición correlativa de un proyecto de vida que engloba el proyecto profesional, es decir, la profesión a ejercer que contribuya al desarrollo completo de la persona (Fournet et al. en Echeverría, 1997: 87).

Así, el objetivo fundamental de la intervención orientadora sería «dotar a las personas de las competencias necesarias para poder identificar, elegir y/o reconducir alternativas formativas y profesionales, acordes a su potencial y proyecto vital,

en contraste con las ofertadas por los entornos formativos y laborales» (Echeverría, 1997: 97). De acuerdo con esta afirmación, su autor concreta que las estrategias orientadoras tienen que ir encaminadas tanto a la prevención y a la intervención social como al desarrollo, clarificando valores, desarrollando la capacidad de tomar decisiones y facilitando el desarrollo personal, la socialización y el conocimiento del mundo del trabajo.

A continuación, presentamos la aplicación y evaluación del Programa de orientación para la inserción laboral destinado a los internos en los centros penitenciarios, integrado en el currículum de los cursos de formación profesional y/o ocupacional que se llevan a cabo en estos centros.

2. ANÁLISIS DE LA SITUACIÓN INICIAL

Nuestra experiencia en los centros penitenciarios nos ha permitido detectar déficits en la formación y orientación para la inserción de los internos en el mercado laboral. Si bien es cierto que, en un primer momento, la intervención está encaminada a mejorar la cualificación profesional con todo el proceso que conlleva de toma de decisiones, clarificación de valores, etc. nos damos cuenta que posteriormente también tenemos que intervenir para analizar y mejorar las competencias individuales facilitando los conocimientos y habilidades necesarias, así como aumentando la autoestima profesional para que los internos puedan desenvolverse eficazmente en el mercado laboral a la hora de buscar un trabajo. Los internos presentan un nivel bajo de motivación hacia la búsqueda activa de trabajo y una falta de preparación en cuanto a las estrategias y técnicas necesarias para llevar a cabo esta búsqueda.

Valverde (1991) lo resume cuando afirma que ni su aspecto, ni su lenguaje son los más adecuados para inspirar confianza a la hora de buscar un trabajo, y constata la necesidad de intervenir en este sentido:

«En efecto aunque logremos que el recluso alcance una formación y especialización laboral, si no le dotamos de unas habilidades sociales mínimas y de las técnicas necesarias para enfrentarse a la búsqueda de un empleo cuando salga en libertad, no habremos avanzado nada en su proceso de adaptación social» (Valverde, 1991: 199).

3. PROGRAMA DE ORIENTACIÓN PARA LA INSERCIÓN LABORAL

Son numerosas las investigaciones que se han llevado a cabo sobre la inserción laboral de diversos colectivos (jóvenes, poblaciones con necesidades educativas especiales, mujeres, etc.) y desde diferentes perspectivas, tanto socio-económica, sociológica, como psicológica, todas ellas dirigidas hacia la descripción del fenómeno de la inserción social y profesional en sus características definitorias y de las variables implicadas en la explicación del proceso (Figuera, 1995).

Si nos centramos en el marco de la orientación, Isus (1996) apunta cinco enfoques de orientación en el ámbito laboral: el de ocupabilidad, el de selección y ajuste, el centrado en el desarrollo de las organizaciones, el de desarrollo de la carrera y para el desarrollo local y regional.

Lo que aquí centra nuestra atención son los enfoques de ocupabilidad, pues describen el proceso de inserción desde la perspectiva puntual del logro de ocupación. Además integran tanto factores personales como sociales en la explicación del modelo.

Dentro de los enfoques de ocupabilidad encontramos el modelo psicobiológico de empleabilidad, el balance de competencias y el modelo de inserción laboral. El modelo psicobiológico se centra en el análisis de los efectos psicológicos que provoca la situación de desempleo. El balance de competencias tiene el objetivo de acompañar en la autoevalaución de las competencias personales, profesionales aptitudinales y motivacionales para elaborar un proyecto profesional e incluso de formación que garantice el futuro. Por último, el modelo de inserción pretende analizar las variables que pueden aumentar la empleabilidad de las personas.

Para llevar a término nuestra investigación nos hemos basado en este último modelo. La mayoría de autores que trabajan el tema de la inserción laboral (Blanch, 1990; Casals et al., 1989, 1991; Esteve, 1999; Figuera, 1996; Filella, 1998; GAIU,1994; Masjuan et al., 1995, 1996; Montané, 1993; Pérez Escoda, 1996; Villar, 1997, entre otros) están de acuerdo en agrupar las variables que intervienen en la inserción laboral en tres bloques:

- el mercado laboral o determinantes situacionales
- los determinantes personales
- los determinantes de inserción

En concreto se fundamenta en el *modelo de orientación ocupacional de Montané* (1991). Este modelo se basa en la interacción dinámica de tres elementos: *el mercado laboral, el individuo que busca trabajo y las estrategias para la búsqueda de trabajo.*

En cuanto al *mercado laboral,* la relación entre las ofertas y demandas de empleo es el primer aspecto que hemos de evaluar, ya que la aproximación a las posibilidades reales de inserción viene dada por la relación entre las ofertas de colocación y el número de solicitantes por cada tipo de ocupación.

Por otro lado, *el individuo demandante de empleo* está condicionado por su currículum, por su capacidad de aprendizaje teórico-técnico y por las actitudes de inserción (autoestima profesional, disponibilidad laboral y madurez ocupacional).

Finalmente, la probabilidad de éxito en el proceso de búsqueda de empleo dependerá del dominio *de las estrategias de inserción y de la capacidad para ponerlas en práctica* (elaboración de la carta de presentación y del currículum, planificación de la búsqueda de ofertas, responder a una entrevista de selección de personal, pruebas psicotécnicas,...). Jariot y Martínez (1994: 241) corroboran este enfoque cuando apuntan las tres variables más influyentes a la hora de encontrar trabajo: en primer lugar la variable «saber comportarse en una entrevista», seguida de la variable «tener confianza en encontrar empleo» y en tercer lugar «estar dispuesto a aprender cosas nuevas de la profesión».

En este sentido, los contenidos del programa que presentamos se agrupan en cuatro bloques:

I. Análisis del mercado laboral y del rol profesional (necesidades y competencias profesionales).

II. Cómo buscar un lugar de trabajo en: OTG, ETT, consultores de empresa, medios de comunicación, amigos y conocidos, servicios de inserción de las comunidades autónomas, ayuntamientos y otras instituciones,...).

III. Proceso de selección (carta de solicitud, currículum vitae, entrevista, pruebas psicotécnicas...).

IV. Derechos y deberes del trabajador y modalidades de contratación.

Cada unidad esta compuesta de cuatro apartados:

I. Ficha de programación de la unidad, donde consta:
   — la finalidad de la unidad
   — el esquema de los contenidos
   — la metodología y las actividades
   — los recursos necesarios para llevarla a cabo
   — la evaluación

II. Contenidos de la unidad: en este apartado se exponen los contenidos que el profesor transmitirá a los alumnos.

III. Fichas didácticas de la unidad: tienen el objetivo de facilitar el proceso de enseñanza-aprendizaje a los alumnos. El número varía en función de los contenidos de cada unidad.

IV. Ficha Resumen: la deben cumplimentar los alumnos y sirve como feed-back del proceso de enseñanza-aprendizaje (Filella, 1998).

Aparte de los materiales que se sugieren en las fichas, el profesor puede utilizar cualquier tipo de material que considere adecuado para trabajar la unidad.

4. ESTRATEGIAS PARA LA APLICACIÓN DEL PROGRAMA

La estrategia para aplicar el programa coincide con *el módulo integrado en el currículum de formación* propuesto por Montané (1990), en el cual los programas de orientación para la inserción laboral se incluyen dentro de los cursos de formación profesional y/o ocupacional, y son los profesores de dichos cursos los que aplican el programa con el asesoramiento y apoyo de un orientador laboral.

En nuestro caso el programa se incluye dentro de los cursos de formación profesional y ocupacional que se llevan a cabo en los centros penitenciarios, es decir, se van incorporando actividades del programa de orientación para la búsqueda de empleo a lo largo del curso:

- El profesor del curso asume la responsabilidad de preparar profesionalmente a los alumnos transmitiendo los contenidos propios del curso (electricidad, construcción, confección, carpintería, automoción,...) al tiempo que facilita las técnicas y habilidades necesarias para la búsqueda de trabajo.
- El orientador laboral (normalmente el pedagogo en los centros penitenciarios) forma a los profesores de los cursos (modelo de consulta (programa de formación de formadores) y desarrolla un papel de apoyo y coordinación de todo el proceso.

Dada la movilidad dentro de los grupos de formación en los centros penitenciarios sería conveniente que la aplicación de este módulo siga una temporalización continua,

es decir, que se procure dar un hilo conductor a las diferentes actividades que se reparten gradualmente a lo largo del curso.

5. EVALUACIÓN DEL PROGRAMA

Son muchos los modelos de evaluación de programas existentes y que recogen diferentes autores Colás y Rebollo, 1993; Medina y Villar, 1996; Latorre et al., 1996; Tejedor et al., 1994, entre otros. Castillo y Gento (1996) proponen una clasificación de los modelos agrupados en tres tipos: el conductista-eficientista, el de corte humanístico y el holístico.

Dentro del modelo conductista-eficiencista, enconctramos el modelo de consecución de objetivos de Tyler, el de la figura de Stake, el de planificación educativa de Cronbach, el CIPP de Stufflebeam, el CSE de Alkin, estos tres últimos, también llamados modelos de facilitación de decisiones o de toma de decisiones.

El modelo humanístico acoge el de contraposición de Owens y Wolf, el de atención al cliente de Scriven y el de crítica artística de Eisner y por último, dentro del modelo holístico se incluyen la evaluación respondente de Stake, la evaluación iluminativa de Parlett y Hamilton y la evaluación holística de Macdonald (Castillo y Gento en Medina y Villar, 1996).

En nuestro estudio hemos optado por el modelo CIPP de Stufflebeam para evaluar la aplicación de nuestro programa, utilizado con éxito en varias investigaciones evaluativas (Martínez Muñoz, 1995; Pérez Escoda, 1996; Pineda, 1994, entre otros). Tal como afirma el autor «la evaluación es un instrumento para ayudar a que los programas sean mejores para la gente a la que deben servir» (Stufflebeam en Bartolomé, 1990). Este modelo propone cuatro dimensiones evaluadoras: Contexto, Entrada, Proceso y Producto, las cuales corresponden a cuatro tipos de decisiones, respectivamente: Planificación, Programación, Aplicación y Reciclaje.

A grandes rasgos podemos afirmar que:

— La evaluación de contexto juzga si los objetivos propuestos son suficientemente coherentes respecto a las necesidades valoradas.
— La evaluación de entrada tiene como objetivo valorar el diseño y los recursos disponibles para conseguir las finalidades del programa.
— La evaluación de proceso consiste en una continua verificación de la aplicación del programa. Aporta la información necesaria para saber como se desarrolla.
— La evaluación de producto se lleva a cabo una vez finalizado el programa y valora la eficacia, la efectividad, la eficiencia y el impacto (Stufflebeam, 1987).

Una vez seleccionado el modelo de evaluación es necesario establecer el diseño de la evaluación. Actualmente la tendencia dominante en las investigaciones evaluadoras es buscar la complementariedad metodológica, porqué la pluralidad permite ampliar el conocimiento de la realidad.

Considerando las metodologías cuantitativas podemos citar la experimental, la cuasiexperiemntal y la no experimental o ex post-facto. La metodología experimental y la cuasiexperimental buscan las relaciones de causalidad a partir de la manipulación intencionada de la variable independiente para descubrir los efectos que produce esta manipulación sobre la variable dependiente, a diferencia de los diseños no experimen-

tales o ex post-facto donde los efectos de la variable independiente ya se han producido y básicamente podemos estudiar sus resultados, describiendo, correlacionando, comparando. En los diseños cuasiexperimentales se parte de situaciones reales o de campo, este es el caso de la mayoría de investigaciones en educación (Pérez Escoda, 1996: 97), donde el evaluador trabaja con grupos constituidos y no puede escoger la muestra al azar, como es el caso de los diseños experimentales.

Nuestra investigación, también, queda enmarcada en el campo de la educación y, como es habitual, no queda inicialmente garantizada la equivalencia entre los grupos experimentales y los grupos control; por lo cual hemos optado por un diseño de la metodología cuasiexperimental, en concreto el de grupos no equivalentes pretest-postest con grupo control.

Como técnicas para llevar a cabo la investigación hemos utilizado un cuestionario de inserción laboral que ha actuado, a la vez, como pretest y postest, y varios cuestionarios para conocer la opinión de las personas implicadas (profesores y alumnos) en el programa de carácter más cualitativo, que complementan la metodología cuantitativa (Pérez Escoda en Filella, 1998).

Evaluación de contexto

La muestra utilizada ha sido de 256 internos que asistían a cursos de formación ocupacional en el Centro penitenciario Ponent en Lleida, entre octubre de 1993 y julio de 1996, de los cuales 189 formaban el grupo experimental y 67 el grupo control, es decir, estos últimos no participaban en la aplicación del programa. Las especialidades de los cursos fueron Automoción, Artes Gráficas, Confección Industrial, Construcción, Electricidad, Encuadernación, Carpintería e Informática.

La muestra estaba formada por internos con una condena inferior a 4 años, de los cuales, un 11% eran mujeres y un 89% hombres, entre 25 y 35 años, con un bajo nivel de estudios (Certificado de Escolaridad) y que carecian de conocimientos sobre la búsqueda de trabajo. Estas características, antes mencionadas, las encontramos representadas tanto en el grupo experimental como en el grupo control, ya que el nivel de inserción inicial no presenta diferencias significativas.

Evaluación de entrada

Para adecuar mejor el programa, al contexto descrito, pedimos la colaboración de un grupo de expertos o jueces. Una vez analizado estadísticamente el resultado mediante la prueba del Coeficiente de Concordancia de «W» de Kendall consideramos satisfactorio el resultado obtenido de 0,61 para continuar la investigación.

Evaluación de proceso

En relación a la evaluación de proceso nos servimos de las valoraciones emitidas por todos los profesores de los cursos de formación ocupacional y por un grupo de 52 alumnos (escogidos aleatoriamente) que participaron en el programa.

En general podemos afirmar que la aplicación del programa no comportó grandes dificultades en la mayoría de los cursos. En cualquier caso se contó con la colaboración del orientador laboral del centro, siempre que el profesor lo consideró oportuno.

De la información extraída hemos de destacar que los profesores dedicaron una media de aproximadamente 8 horas a la preparación del material para impartir el módulo de inserción. La media de horas dedicadas a la aplicación fue de 18 h. 30 m., temporalización que se considera adecuada, entendiendo que el módulo propuesto, por el Departamento de Trabajo, para los cursos de formación ocupacional es de 20 horas. Si analizamos el tiempo de forma global, se observa que se ha incidido más en la parte práctica, pues el número de horas prácticas dobla el número de horas teóricas, aspecto que valoramos positivamente.

En cuanto a la valoración de los alumnos, hay que considerar que en un principio no aceptaron con mucho interés el programa, pero esta desmotivación inicial se fue superando a medida que avanzaba el curso. La opinión de los alumnos manifestada en el cuestionario anónimo de valoración final fue muy positiva.

En referencia a los ítems sobre la utilidad de incluir en el curso este tipo de actividades, el 67% de los internos encuestados contestaron «Mucho» y el resto «Bastante». Estas respuestas coinciden con el siguiente ítem sobre la consideración de sí ha valido la pena participar en estas actividades, donde las 2/3 partes han contestado que «Mucho» (67,3%) y el resto «Bastante».

Evaluación de resultados

En este apartado medimos los resultados que se han obtenido, teniendo en cuenta la eficacia, la efectividad, la eficiencia y el impacto.

Como ya hemos anunciado, utilizamos una metodología de diseño cuasiexperimental, pretest-postest con grupo control.

Grupo Experimental	Pretest ..... PROGRAMA ..... Postest
Grupo Control	Pretest ..... Postest

A partir de este diseño el análisis estadístico se basa en tres situaciones diferentes. Tendremos en cuenta si la situación es de comparación o de relación, si se trata de grupos independientes o relacionados y si nos encontramos delante de datos de distribución normal (situación paramétrica) o de distribución libre (situación no paramétrica):

- Análisis de la situación inicial. Comprobamos la situación de partida de los grupos experimental y control mediante la comparación de medias para grupos independientes, partiendo de los resultados del pretest (cuestionario de inserción laboral): «t» Student o «U» de Mann-witney.
- Análisis de la situación final. Comparamos las medias para grupos independientes a partir de los resultados del postest: «t» Student o «U» de Mann-witney.

- Análisis entre situación inicial y final en cada grupo. Utilizamos la comparación de medias para grupos relacionados: «t» Student o «T» Wilcoxon.

a) En relación con la eficacia

La comparación entre las medias de los grupos experimental y control en el pretest evidencia la equivalencia en la situación inicial. En cambio, si comparamos las medias de los dos grupos obtenidas a partir del postest, es decir, en la situación final, podemos observar que sí que existen diferencias con un nivel de significación de 0,001, a favor del grupo experimental (ver tabla 1).

Finalmente, si comparamos las situaciones inicial y final de cada grupo por separado, podemos ver que las diferencias que encontramos en el grupo control no son significativas estadísticamente; mientras que la media global del grupo experimental ha aumentado significativamente con un nivel de probabilidad asociada de riesgo inferior al uno por mil. A partir de aquí podemos atribuir a la aplicación del programa la mejora en el nivel de inserción laboral de los alumnos del grupo experimental, es decir, han aumentado sus conocimientos y habilidades, así como su autoestima profesional.

TABLA 1
CONTRATE DE LAS MEDIAS OBSERVADAS ENTREGRUPOS E INTRAGRUPOS EN LAS SITUACIONES INICIAL Y FINAL

	Grupo experimental	Grupo Control	
0 Pretest	5,52	5,54	n. s.
0 Postest	7,68	5,55	* * *
* * * p<0,001	* * *	n. s.	

Si comparamos las medias de todas las variables que intervienen hemos de decir que las mejoras se producen independientemente de estas variables y de su interacción. Las variables que hemos considerado son el sexo, la edad, la condena, el nivel de estudios, los cursos de formación ocupacional realizados anteriormente, la experiencia laboral, los conocimientos y las habilidades previas para la búsqueda de trabajo.

Los contrastes realizados hasta ahora evidencian una influencia positiva de nuestro programa, que se corroboran cuando controlamos los efectos del pretest: a través del coeficiente de Correlación Múltiple (0,696) y su coeficiente de Determinación (0,834), observamos como el 83,4% de la varianza obtenida se explica por la variable independiente (programa).

Estos datos estadísticos se complementan con la valoración de los profesores e internos que han participado en la aplicación del programa, teniendo en cuenta la siguiente escala: Mucho, Bastante, Poco, Nada.

• **Los profesores**

La valoración que hacen éstos del aprendizaje del programa es muy positiva. El total del profesorado (8) creen que los alumnos han aumentado «Mucho» los conocimientos sobre inserción laboral y 7 de ellos consideran que también han mejorado «Mucho» en el desarrollo de las habilidades de inserción; sólo uno contesta «Bastante» en este último ítem.

En cuanto a la autoestima profesional de los internos, 7 de los profesores consideran que ha aumentado «Bastante» y uno considera que «Poco». De todas maneras este último ítem es difícil de valorar.

• **Los alumnos**

La valoración de los alumnos respecto a estas actividades también es muy favorable:

En respuesta a la pregunta sobre la mejora de la información relacionada con el mercado de trabajo encontramos un 33% de los internos que considera que ha mejorado «Mucho», un 63% que ha mejorado «Bastante» y un 4% que su información ha mejorado «Poco».

En cuanto a las habilidades de búsqueda de trabajo, los resultados son los siguientes: un 29% de los internos considera que están «Muy» capacitados, un 67% «Bastante» y únicamente un 4% considera que no está más capacitado para encontrar trabajo de lo que lo estaba antes de comenzar el programa.

Respecto a la motivación para la búsqueda de trabajo, un 62% ha contestado «Mucho», un 36% «Bastante» y un 2% que han aumentado «Poco» las ganas de buscar trabajo.

Las respuestas sobre la aceptación de nuevos cursos de formación han sido: un 52% lo haría de forma «Muy» positiva, un 42% ha contestado «Bastante» y un 12% «Poco».

A partir de los datos que hasta ahora hemos ido apuntando, podemos considerar eficaz el programa porqué ha aumentado el nivel de inserción (conocimientos, habilidades y autoestima profesional) de los internos que han participado en él, independientemente de las otras variables que han intervenido y que hemos estudiado. Consideramos fundamental la variable autoestima profesional en el caso de la población interna en los centros penitenciarios.

b) En relación con la efectividad

Para valorar la efectividad nos centraremos en la utilidad real del programa, es decir, tratamos, seis meses después de finalizar los cursos, de comprobar la inserción laboral de los internos que participaron en el programa, comparando los índices de inserción conseguidos con los del grupo control. Presentamos los resultados obtenidos en la siguiente tabla:

TABLA 2
COMPARACIÓN DE LA INSERCIÓN LABORAL ENTRE EL GRUPO EXPERIMENTAL Y EL GRUPO CONTROL

GRUPO	EXPERIMENTAL		CONTROL	
Centros penitenciarios	102	70,3%	30	58,8%
Inserción laboral	21	14,4%	5	9,80%
No trabajan	10	6,89%	3	5,88%
No localizados	4	2,75%	7	13,7%
Residencia en el extranjero	4	2,75%	3	5,88%
Defunción	3	2,06%	3	5,88%
Realizan estudios FO	1	0,68%	0	0%
Total	145	100%	51	100%

Para valorar los resultados hemos de tener en cuenta que únicamente se ha podido comprobar la inserción laboral del 30% de los internos del grupo experimental y del 42% del grupo control; el resto todavía cumplía la condena.

Entendiendo que la muestra de los internos en libertad es muy reducida, desde nuestro punto de vista los resultados obtenidos pueden considerarse satisfactorios, ya que el 48% de los sujetos del grupo experimental que están en libertad ha encontrado trabajo y el 23% del grupo control también. Hay que destacar que muchos de los internos ocupados no desarrollan trabajos relacionados con la especialidad del curso que realizaron.

TABLA 3
COMPARACIÓN DE LA INSERCIÓN LABORAL SOBRE EL TOTAL DE INTERNOS EN LIBERTAD ENTRE EL GRUPO EXPERIMENTAL Y EL GRUPO CONTROL

GRUPO	EXPERIMENTAL		CONTROL	
Total en libertad	43	100%	21	100%
Inserción laboral	21	48,8%	5	23,8%

c) En relación con la eficiencia

Podemos constatar que los beneficios obtenidos rentabilizan la inversión realizada en recursos funcionales, humanos y materiales, puesto que la aplicación del programa no supone ningún incremento excesivo en el trabajo de los profesores.

d) En relación con el impacto

Una vez acabada la investigación, no destacamos consecuencias no previstas debidas a la aplicación del programa. Consideramos que el hecho de estar en contacto con todos los implicados a lo largo de toda la experiencia ha reducido al máximo las situaciones imprevistas.

El programa que hemos evaluado supone el *primero* de los tres niveles progresivos de orientación ocupacional que proponemos, en los cuales se establece y coordina un itinerario formativo-ocupacional con el objetivo final de insertar laboralmente al interno cuando obtenga la libertad condicional i/o definitiva:

Segundo nivel

Este nivel se basa en el módulo formativo y se aplica a aquellos internos que disfrutan de permisos ordinarios de salida. La acción formativa es más puntual y especializada, y se incide más en la aplicación práctica de los contenidos, ya que los internos tienen la posibilidad de aplicar los conocimientos y habilidades durante las salidas al exterior. La intervención es grupal y es llevada a cabo por el orientador laboral (el pedagogo del centro), que dinamiza el grupo, favoreciendo el debate entre los alumnos a lo largo de 20 .. 30 horas distribuidas en sesiones de dos horas. La evaluación se realiza una vez han regresado de permiso.

Tercer nivel

Podemos distinguir dos casos en función de la situación penitenciaria del usuario:

a) *Internos en tercer grado* de tratamiento penitenciario que se encuentran ubicados en secciones o centros abiertos y a quienes se les permite salir diariamente a la comunidad para formarse, buscar empleo o trabajar. En este caso la intervención la realiza el educador asignado por el centro que actúa como orientador e «insertor» laboral, facilitando *individualmente* las técnicas y habilidades.
b) *Liberados condicionales y liberados definitivos,* cuyo seguimiento se realiza por las comisiones territoriales de asistencia social penitenciaria (CTASP) de cada provincia. La orientación la lleva a cabo el «insertor» laboral de dicha comisión.

Es muy importante que el paso a este último eslabón se produzca de forma coordinada y progresiva, estableciendo un circuito de derivación de casos desde el Centro Penitenciario con un mínimo de tres meses de antelación a la obtención de una plaza en el centro abierto correspondiente o a la puesta en libertad.

La intervención se concretará en tres líneas básicas de actuación:

— Asesoramiento e información puntual: atención de demandas puntuales en las que no se realiza seguimiento.
— Formativa: previa a la inserción laboral en los casos de usuarios que carecen de formación profesional i/o experiencia laboral.
— Inserción laboral directa, dentro de la cual, se contemplan dos opciones:

- Orientación para la inserción laboral: el usuario cuenta con experiencia laboral (que podría haber adquirido en el centro penitenciario, o bien, antes del ingreso), pero no posee los conocimientos necesarios para buscar empleo, así que se ha de proporcionar al usuario las técnicas y habilidades de inserción necesarias.
- Búsqueda individualizada: en este caso el usuario cuenta con una buena experiencia laboral reciente, o bien, ha participado en todos los niveles anteriores de inserción durante su estancia en la institución y puede buscar trabajo directamente.

Hay que destacar que es posible, y a la vez necesaria, la coordinación entre los diferentes niveles.

BIBLIOGRAFÍA

BARTOLOMÉ, M. (1990). Evaluación y optimización de los diseños de intervención. *Revista de Investigación Educativa*, 8, 16, 39-61.

BLANCH, J.M. (1990). *Del viejo al nuevo paro*. Barcelona: PPU.

BLANCO, A. I. (1991). *Bases para el establecimiento de un modelo de intervención diferencial en el campo de la delincuencia juvenil: una aproximación biográfica*. Tesis doctoral. Madrid: Universidad Complutense de Madrid.

CANTOR, D. y LAND, K. (1991). Exploring possible temporal relationships of unemployment and crime: a comment on Hale and Sabbagh. *Journal of Research in Crime and Delinquency*, 28, 4.

CASAL, J.; MASJUAN, J.M. y PLANAS, J. (1989). *La inserción social y profesional de los jóvenes*. Barcelona: ICE-UAB y CIDE.

CASAL, J.; MASJUAN, J.M. y PLANAS, J. (1989). «Educación e inserción social de los hombres y mujeres de 25 años». Revista de Educación, 290, 177-196.

CASAL, J.; MASJUAN, J.M. y PLANAS, J. (1991). *La inserción social y profesional de los jóvenes*. Barcelona: ICE-UAB.

COLÁS, M.P. y REBOLLO, M.A. (1993). *Evaluación de Programas. Una guía práctica*. Sevilla: Kronos.

CÓRCOLES, A. (1994). Els efectes de la desocupació. *Papers d'Educació d'Adults*, 21, 16-19.

ECHEVERRÍA, B. (1997). Inserción sociolaboral. *Revista de Investigación Educativa*, 15, 2, 85-115.

ESTEVE, M. (1999). *Variables de inserción de los titulados de la universidad de Lleida*. Tesis de licenciatura. Universidad de Lleida.

FARRIGTON, D.P. y WEST, D. (1990). *The Cambridge study in delinquent development: a long-term follow of 411 London males*. Cambridge.

FIGUERA, P. (1995). Panorámica de la investigación sobre los procesos de inserción socio-profesional en *Revista de Investigación Educativa*, 25, 125-148.

FIGUERA, P. (1996). *La inserción socioprofesional del universitario en el mercado laboral*. Barcelona: EUB.

FILELLA, G. (1998). *L'orientació ocupacional en els centres penitenciaris*. Lleida: Pagès.

GAIU (1994). *Informe de l'Enquesta d'opinió als graduats de la UB: Promoció 1993*. Barcelona: Universidad de Barcelona (papel).

ISUS, S. (1993). *La inserción académica de los alumnos de curso de orientación universitaria*. Tesis doctoral. Barcelona: Universidad de Barcelona.

ISUS, S. (1996). *Proyecto docente*. Universidad de Lleida.

JARIOT, M. y MARTÍNEZ, M. (1994). Dimensiones implicadas en la inserción laboral análisis a partir de una escala. *Revista de Investigación Educativa*, 23, 240-244.

LATORRE, A.; DEL RINCÓN, D. y ARNAL, J. (1996). *Bases metodológicas de la investigación educativa*. Barcelona: GR92.

MARTÍNEZ MUÑOZ, M. (1995). *La orientación del clima en el aula*. Tesis doctoral. Universidad Autónoma de Barcelona.

MASJUAN, J.M. TROIANO, H.; VIVAS, J. y ZALDIVAR, M. (1995). *Estudi dels itineraris de formació i ocupació dels nous titulats universitaris*. Bellaterra: ICE-UAB.

MASJUAN, J.M.; TROIANO, H.; VIVAS, J. y ZALDIVAR, M. (1996). *La inserció professional dels nous titulats universitaris*. Barcelona: ICE-Universitat Autónomo de Barcelona.

MONTANÉ, J. (1990). *La orientación profesional*. Barcelona: Fondo de Formación.

MONTANÉ, J. (1991). Cómo tratar la inserción laboral. *Herramientas*, 14, 36-41.

MONTANÉ, J. (1993). *Orientación ocupacional*. Barcelona: CEAC.

MORRIS, N. (1981). *El futuro de las prisiones*. 2ª edición. Méjico: XXI.

PÉREZ ESCODA, N. (1996) *Programa de formació per a la inserció laboral*. Barcelona: Ceac.

PINEDA, P. (1994). *La formació a l'empresa. Planificació i avaluació*. Barcelona: Ceac.

REDONDO, S. (1992). Educación en prisiones y reinserción social. *En XVIII Reunión científica de la Asociación Española para la Educación Especial*. València: Institut Valencià de Serveis Socials.

STUFFLEBEAM, D.L. y SHINKFIELD, A.J. (1987). *Evaluación sistemática*. Barcelona: Paidós.

SOOTHILL, K. y HOLMES, J. (1981). Finding employment for ex-prisoners: a ten-year follow-up study. *Howard Journal of Penology and Crime Prevention*, 20, 1, 29-36.

TEJEDOR, F. J., GARCÍA-VALCÁRCEL, A. y RODRÍGUEZ CONDE, Mª J. (1994). Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo. *Revista de Investigación Educativa*, 23, 93-129.

VALVERDE, J. (1991). *La cárcel y sus consecuencias*. Madrid: Popular.

VILLAR, E.; JUAN, J.; COROMINES, E. y CAPELL, D. (1997). *Evaluación de los cursos de acción previos al proceso de inserción laboral. Un estudio piloto en la universidad de Girona*. Sevilla: AIDIPE.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 97-118

MODELIZACIÓN ESTOCÁSTICA DE LA ELECCIÓN DE CARRERA UNIVERSITARIA Y DE LOS RESULTADOS OBTENIDOS EN PRIMER CURSO DE LAS TITULACIONES DE PEDAGOGÍA Y MAGISTERIO

Gómez García J., Hernández Pina F., Buendía Moya F., Carmona Martínez M.
Universidad de Murcia

RESUMEN

Este trabajo presenta modelizaciones logit-binomial de la «demanda de Educación Superior» y de los «resultados obtenidos en primer curso» de los graduados en Enseñanzas Medias que se matriculan en la Facultad de Educación de la Universidad de Murcia en las titulaciones de Magisterio y Pedagogía. Utilizando dichos modelos se intenta establecer los condicionantes de la elección de alternativas por parte de los estudiantes. Incorporamos, entre otras, aquellas variables que la «teoría del capital humano» considera como relevantes en las decisiones de invertir en educación. Los modelos estimados explican: por un lado, las elecciones educativas de los estudiantes en función de los logros educativos de sus madres, el expediente académico y la opción de COU; y por otro, los resultados obtenidos por los estudiantes en primer curso observando que las calificaciones previas y la Opción de COU son las variables que mejor explican el éxito o el fracaso.

ABSTRACT

This research presents the logit-binomial modelling of the demand for High Education and of the secondary first year students´ results, who have registered at the Education Faculty at the University of Murcia. Using such models, it has been tried to establish the conditioning factors which influence the students´s election of course options. We introduce, among others,

those variables the «human capital theory» considers most relevant to the decisions for investing in Education. The estimated models explain, in one hand, the students´ educational choices considering their mothers´s educational success, their academic performance and COU option students choose; on the other hand, these models explain the results students obtain in their first year at university, taking into account that their previous results and the COU option are the variables which best explain and reflect the students´ success or failure.

1. INTRODUCCIÓN

«La evaluación de la calidad de la educación universitaria se ha convertido no sólo en una prioridad sino también en una exigencia tanto para las propias universidades como para los gobiernos y las administraciones públicas» (MEC, 1998).

Con esta frase se abre el documento Guía de Evaluación del Plan Nacional de la Calidad de las Universidades. De sobra es conocida la prioridad que la evaluación institucional está recibiendo por parte de las universidades de todo el Estado. Desde que en 1995 se estableciese el primer Plan de Evaluación cada año son más las universidades que vienen tomando parte en las convocatorias de ayudas a proyectos de evaluación del MEC con el fin de someter a evaluación un número cada vez más creciente de titulaciones.

Con el fin de facilitar la tarea evaluativa, el MEC ha designado un Comité Técnico encargado de elaborar una Guía de Evaluación. En dicha Guía aparecen las normas y los protocolos que permitan a las universidades seguir procesos y metodologías similares. Ante la proliferación de formas, tipos y procedimientos de evaluación, el MEC ha optado por un modelo globalizado en el que se articule la evaluación interna (autoevaluación) y la evaluación externa, así como las evaluaciones formativa y sumativa. Esta evaluación cubre las tres grandes dimensiones de la universidad: la enseñanza, la investigación y los servicios.

En la autoevaluación de la enseñanza, la Guía determina los ámbitos en los que se deberá centrar dicha evaluación y el tipo de datos a recoger. Uno de dichos ámbitos es el referente al contexto de la universidad y, por consiguiente, al contexto de la titulación.

En la Guía observamos apartados que van desde el contexto de la universidad, la evolución de la titulación objeto de evaluación, las metas y objetivos; el programa de formación; el desarrollo de la enseñanza, hasta los resultados de tal enseñanza.

La investigación sobre la calidad en la universidad (Apodaca y Lobato, 1997), la búsqueda de indicadores en la universidad (MEC, 1999) y los informes de Plan Nacional de Evaluación de la calidad de las Universidades ponen de manifiesto la necesidad de generar un corpus de información válida y fiable que ayude en cualquier toma de decisiones. La primeras fases del plan han permitido recoger un gran número de datos que, sin duda, han llevado a conclusiones importantes. Sin embargo, estos datos no bastan, sino que hay que orientar el trabajo hacia la búsqueda de indicadores, tarea que parece haberse impuesto el Consejo de Universidades al constituir un grupo de trabajo para la definición de un catálogo de indicadores.

Centrándonos en una de las áreas del Plan Nacional, la evaluación de la enseñanza ha evitado reducirse a la evaluación del profesor. Por el contrario, el modelo asumido por el MEC ha tratado de incluir una variedad de ámbitos sobre los que recoger datos amplios y diversificados. Una de las dimensiones sobre las que se solicitan datos en la Guía, son los referidos a la titulación, centrándose fundamentalmente en la oferta de plazas, la demanda de la titulación y la matrícula (calificaciones de acceso, nota corte, admisión en primera opción, calidad de acceso, etc.). Sin menospreciar la importancia de dichos datos de entrada, en este trabajo abordamos otros componentes de entrada que podrían ayudar a valorar cual es el perfil que el alumnado tiene al elegir una determinada titulación y qué peso tienen esas mismas variables en la explicación el rendimiento académico.

Aunque el valor económico de la «educación como un bien» ha sido contemplado en la literatura económica desde los economistas clásicos (Smith, 1776): las personas con niveles más altos en educación ganan más en promedio. La idea de que las actividades que favorecen la productividad de los seres humanos pueden ser tratadas como una forma de inversión, genera el concepto de capital humano que nace con el discurso pronunciado por T.W. Schultz, ante la reunión anual de la American Economic Associatión en 1960.

Con la introducción de la «teoría del capital humano» la relación entre niveles de escolaridad e ingresos futuros ha sido acentuada. La principal aportación de esta teoría es considerar la educación como una inversión. La hipótesis básica de este enfoque es que la educación supone, para el individuo que la adquiere, una acumulación de capital humano que determina un aumento en la productividad de aquel y, por tanto, un incremento en las rentas salariales futuras (Schultz; 1960, 1963, Mincer; 1958, 1974 y Becker; 1964).

Estos beneficios de la educación, ingresos más altos para personas con mayor stock educativo, pueden ser obtenidos por un individuo si él está dispuesto a pagar los costes directos y de oportunidad de la educación. En la línea con la expansión de la teoría del capital humano, muchos estudios han intentado estimar tasas de retorno para la educación (Psacharopoulos, 1980, 1985).

Desde el punto de vista del individuo es este aspecto de la inversión lo realmente relevante, es decir, cómo los ingresos futuros se relacionan con la educación; si embargo, no es relevante para el individuo si los mayores ingresos vienen motivados por un incremento de la productividad o, por el contrario, la educación funciona como un filtro que selecciona personas con cualificaciones específicas (Arrow, 1973; Spence, 1973; Riley, 1979).

Sin embargo, la «teoría del capital humano» ha sido criticada principalmente por contemplar la educación como una inversión y no tener en cuenta aspectos de utilidad o motivos consumo de la educación (Blaug, 1976). El motivo consumo de la educación mide la contribución de la educación a la utilidad, separándola de la dimensión monetaria. Campbell y Siegel (1967) estimaron un modelo consumo de demanda de educación y Michael (1973) y Becker (1964) incorporan el motivo consumo de la educación en un modelo de producción doméstica. Heckman (1976) incluye el motivo consumo de la educación asumiendo que la educación incrementa la eficiencia en la

distribución del ocio. En estos modelos tanto los aspectos monetarios como no monetarios determinan, conjuntamente, la cantidad óptima de educación.

Además de los dos motivos inversión y consumo para demandar educación, hay otro factor importante a incorporar en un modelo de formación de capital humano; nos referimos a las perspectivas futuras de empleo. En general, el desempleo está fuertemente relacionado con el nivel de educación: los individuos con niveles más altos de educación tienen una probabilidad menor de llegar a estar desempleados (Becker, 1964; Ashenfelter y Ham, 1979; Nickell, 1979; Bricall, 2000).

Sin embargo, diversos estudios empíricos han puesto de manifiesto que en la toma de decisiones acerca de la cantidad óptima de educación que se desea recibir, no sólo influyen determinantes económicos sino que también tiene gran importancia el ambiente familiar y la «habilidad escolar» de cada individuo.

En este contexto se sitúa nuestra investigación, que trata de analizar las decisiones de inversión en Educación Superior de los alumnos matriculados en la Facultad de Educación de la Universidad de Murcia, en las siete posibles carreras que este centro ofrece: Licenciatura en Pedagogía (ciclo largo), Maestro en Educación Infantil, Maestro en Educación Primaria, Maestro en Lengua Extranjera, Maestro en Educación Musical y Maestro en Educación Física (ciclo corto). Posteriormente, estudiaremos cual es la evolución de los alumnos que cursan 1º de Pedagogía y de Magisterio, de qué dependen sus resultados en ese curso y si difieren en ambas titulaciones (licenciatura vs diplomatura). Y todo esto dentro del contexto de la Evaluación de la Calidad de la Universidades, es decir queremos contribuir a la determinación de qué otras variables pueden tomarse en consideración para explicar la elección de una titulación y el rendimiento en la misma.

2. METODOLOGÍA

2.1. Muestra

Para llevar a cabo este trabajo se han utilizado los datos de los alumnos matriculados, por primera vez, en primero de todas las titulaciones citadas anteriormente, en el curso 1994-1995. De la muestra se dispone de información sobre las siguientes variables:

1. Nivel de estudios del padre y de la madre, clasificados en «Sin estudios», «Estudios primarios completos», «Bachiller Elemental o asimilados», «Bachiller superior o asimilados», «Diplomatura Universitaria o asimilados» y «Licenciatura, Ingeniería, Arquitectura o asimilados».
2. Situación socioeconómica del alumno, determinada a partir del trabajo del padre y de la madre. Puede tomar las categorías: «Directores o gerentes de empresas de la Administración Pública», «Técnicos o profesionales asociados a titulaciones universitarias o no, de carácter postsecundario», «Empleados administrativos y trabajadores de servicios», «Trabajadores cualificados en agricultura y pesca», «Trabajadores cualificados en la industria, construcción y minería»,

«Trabajadores no cualificados», «Profesionales de las Fuerzas Armadas» y «Personas que no han tenido trabajo remunerado».

3. Opción elegida en COU: A, B, C o D.
4. Género.
5. Nota en cada curso de BUP.
6. Nota en COU.
7. Nota de cada una de las asignaturas cursadas en 1º de carrera, para las convocatorias de Junio y Septiembre.

Como podemos observar, las dos primeras variables son variables exógenas al individuo y de carácter socioeconómico, y el resto se refieren a características personales y académicas.

3. ANÁLISIS DE LOS DATOS

3.1. Análisis descriptivo de la muestra

En primer lugar, analizaremos las características principales de los participantes que forman nuestra muestra, distinguiendo entre los alumnos de la Licenciatura en Pedagogía y los de Magisterios, con el fin de definir el perfil de los estudiantes de ambos tipos de titulaciones.

Si nos fijamos en el nivel de estudios del padre y de la madre, reflejados en las tablas 1 y 2 y en los gráficos 1 y 2, destaca el hecho de que la mayor parte de los participantes se agrupan en la categoría «2», que corresponde a «Estudios Primarios». Además, en la Licenciatura en Pedagogía es superior el porcentaje de alumnos cuyo padre/madre tiene un nivel de estudios correspondiente a «Estudios primarios» (52,3% frente a 42,7%, y 61,9 frente a 48,5%, respectivamente). Para ambas titulaciones es muy bajo el porcentaje de licenciados entre los padres de los estudiantes, mucho más bajo que el porcentaje de padres sin ningún tipo de estudios.

TABLA 1
NIVEL DE ESTUDIOS DEL PADRE

	Frecuencia		Porcentaje	
	Pedagogía	Magisterio	Pedagogía	Magisterio
1. Sin estudios	38	44	15,9	18,4
2. E. Primarios	125	102	52,3	42,7
3. B. Elemental	28	30	11,7	12,6
4. B. Superior	19	25	7,9	10,5
5. Diplomatura	21	24	8,8	10,0
6. Licenciatura	8	14	3,3	5,9
Total	239	239	100,0	100,0

TABLA 2
NIVEL DE ESTUDIOS DE LA MADRE

	Frecuencia		Porcentaje	
	Pedagogía	Magisterio	Pedagogía	Magisterio
1. Sin estudios	47	46	19,7	19,2
2. E. Primarios	148	116	61,9	48,5
3. B. Elemental	22	33	9,2	13,8
4. B. Superior	9	15	3,8	6,3
5. Diplomatura	11	23	4,6	9,6
6. Licenciatura	2	6	0,8	2,5
Total	239	239	100,0	100,0

En cuanto a la variable relativa a la situación socioeconómica familiar del alumno, reflejada en el trabajo del padre y de la madre, no hallamos demasiadas diferencias entre los dos tipos de titulaciones: las ocupaciones que tienen, para ambas carreras, un mayor peso dentro del trabajo de los padres son las de «Empleados administrativos y trabajadores de servicios» (aunque su importancia es relativamente mayor en Magisterio) y «Trabajadores cualificados y operadores de máquinas en la industria, cons-

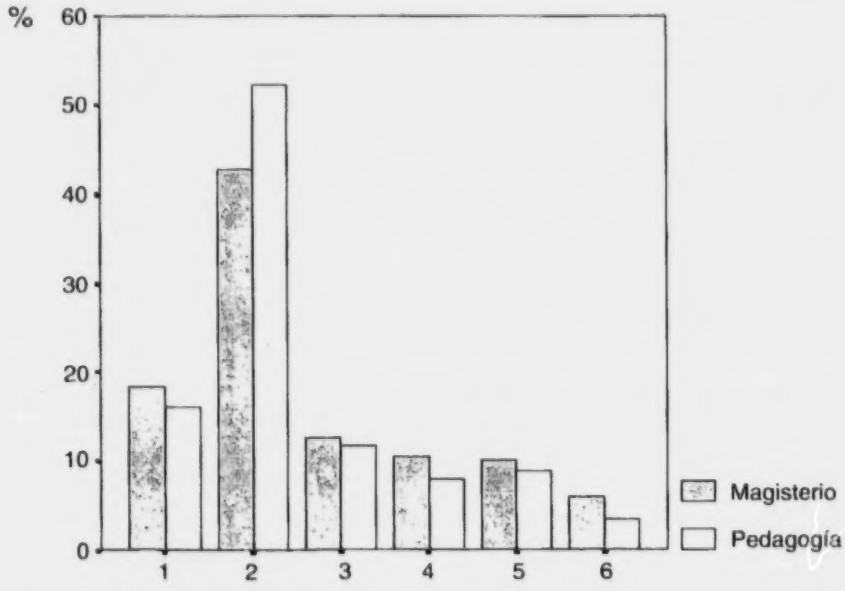


Gráfico 1
Nivel de estudios del padre

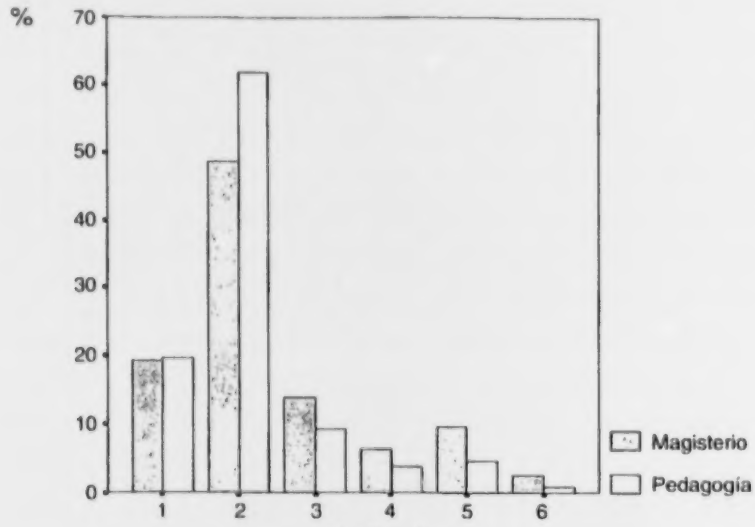


Gráfico 2
Nivel de estudios de la madre

TABLA 3
TRABAJO DEL PADRE

	Frecuencia		Porcentaje	
	Pedagogía	**Magisterio**	**Pedagogía**	**Magisterio**
0. Otros	39	38	16,4	15,9
1. Directores o gerentes de empresas de la Ad. Pública.	3	5	1,3	2,1
2. Profesionales asociados a titulaciones de carácter postsecundario.	27	36	11,3	15,1
3. Empleados administrativos y trabajadores de servicios.	52	61	21,8	25,5
4. Trabajadores cualificados en agricultura y pesca.	13	4	5,5	1,7
5. Trabajadores cualificados en la industria, construcción y minería.	43	49	18,1	20,5
6. Trabajadores no cualificados	40	28	16,8	11,7
7. Prof. de las F. Armadas.	15	13	6,3	5,4
8. Personas que no han tenido trabajo remunerado.	6	5	2,5	2,1
Total	238	239	100,0	100,0

TABLA 4
TRABAJO DE LA MADRE

	Frecuencia		Porcentaje	
	Pedagogía	Magisterio	Pedagogía	Magisterio
0. Otros	7	15	2,9	6,3
1. Directores o gerentes de empresas de la A.P.	1	2	0,4	0,8
2. Técnicos o profesionales asociados a titulaciones de carácter postsecundario.	12	27	5,0	11,3
3. Empleados administrativos y trabajadores de servicios.	19	11	7,9	4,6
4. Trabajadores cualificados en agricultura y pesca.	2	1	0,8	0,4
5. Trabajadores cualificados en la industria, construcción y minería.	4	1	1,7	0,4
6. Trabajad. no cualificados.	23	6	9,6	2,5
7. Profes. de las F. Armadas.	0	0		
8. Personas que no han tenido trabajo remunerado.	171	176	71,5	73,6
Total	239	239	100,0	100,0

* No hay ninguna observación donde esta variable tome el valor 7.

trucción y minería». En la variable relativa a la situación profesional de la madre, destaca el elevado peso del grupo 8 o «Personas que no han tenido trabajo remunerado» en ambos tipos de titulaciones. Estos resultados podemos comprobarlos en las tablas 3 y 4, y en los gráficos 3 y 4.

El comportamiento de la variable «Opción elegida en COU» podemos verlo en la tabla 5. En ambas submuestras, el grupo más numeroso es el de alumnos procedentes de la opción C. Sin embargo, en Magisterio son un 38,5% del total, y en Pedagogía suponen más de la mitad de los alumnos. Respecto a las demás opciones de COU, Magisterio tiene un mayor porcentaje de alumnos en las opciones A y D que Pedagogía 13,9% vs 6,2% y 22,5% vs 14%.

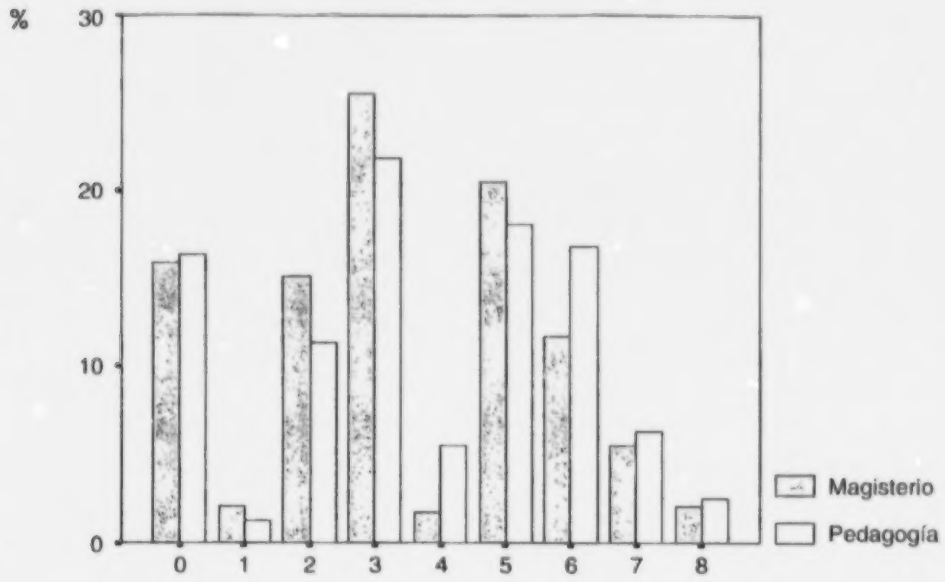


Gráfico 3
Trabajo del padre

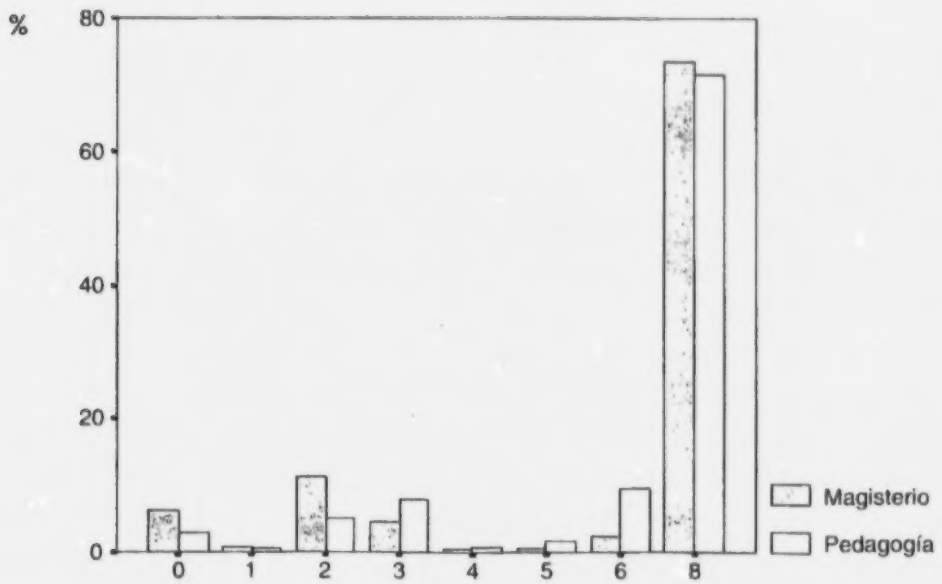


Gráfico 4
Trabajo de la madre

TABLA 5
OPCIÓN ELEGIDA EN COU

		Frecuencia		Porcentaje	
		Pedagogía	Magisterio	Pedagogía	Magisterio
Opción	A	16	34	6,2	13,9
	B	71	61	27,5	25,0
	C	135	94	52,3	38,5
	D	36	55	14,0	22,5
Total		258	244	100,0	100,0

En cuanto a la variable «Género», Pedagogía y Magisterio presentan un porcentaje mayor de mujeres matriculadas en el primer curso de carrera que de hombres, aunque esta tendencia es mucho más acentuada en el caso de la Licenciatura en Pedagogía (tabla 6).

TABLA 6
GÉNERO

	Frecuencia		Porcentaje	
	Pedagogía	Magisterio	Pedagogía	Magisterio
Hombres	38	62	14,7	25,4
Mujeres	220	182	85,3	74,6
Total	258	244	100,0	100,0

Respecto a las notas que los alumnos de la muestra obtuvieron en los cursos de BUP y de COU, destaca el hecho de que la media de cada una de las cuatro variables es superior en el grupo de Magisterio que en el de Pedagogía:

TABLA 7
MEDIA DE LAS NOTAS DE 1º, 2º Y 3º DE BUP Y DE COU

	Pedagogía	Magisterio
Nota 1º BUP	6,3653	7,0334
Nota 2º BUP	6,3387	6,9751
Nota 3º BUP	6,4017	6,9719
Nota COU	6,0090	6,6542

Por último, consideraremos la variable «Resultados», calculada como la media ponderada (por el número de créditos) de la calificación obtenida por cada estudiante en cada asignatura de primer curso (en Junio o Septiembre). Para su cálculo, hemos considerado como «Suspenso» la calificación de las asignaturas a las que el alumno no se ha presentado.

En la tabla 8 podemos comprobar que el porcentaje de alumnos con una nota media en 1º de «Aprobado» y «Notable» es mucho mayor en Magisterio que en Pedagogía:

TABLA 8
RESULTADOS EN 1º DE CARRERA

	Frecuencia		Porcentaje	
	Pedagogía	Magisterio	Pedagogía	Magisterio
(0,1) «Suspenso»	107	37	41,5	15,2
(1,2) «Aprobado»	126	151	48,8	62,1
(2,3) «Notable»	20	54	7,8	22,2
(3,4) «Sobresaliente	5	1	1,9	0,4
Total	258	243	100,0	100,0

Para establecer si existe diferencia significativa entre los resultados en ambas titulaciones hemos efectuado un Análisis de la Varianza (ANOVA). Los resultados han sido: el estadístico F (tabla 9) toma un valor de 64,370 y tiene asociada una probabilidad acumulada de 0,000. Por lo tanto, para un nivel de significatividad del 5%, rechazamos la hipótesis nula de igualdad de medias, y podemos concluir que existen diferencias significativas entre las notas medias obtenidas por los alumnos de 1º de una u otra titulación.

TABLA 9
ANÁLISIS DE LA VARIANZA PARA LA VARIABLE «RESULTADOS» EN FUNCIÓN DEL FACTOR «TITULACIÓN»

Origen de la variabilidad	Suma de cuadrados	Grados de libertad	Media cuadrática	F	Prob
Inter-grupos	26,775	1	26,775	64,370	0,000
Intra-grupos	207,557	499	0,416		
Total	234,332	500			

Como conclusión de este primer apartado, señalaremos que las principales diferencias halladas entre los estudiantes de primer curso de Pedagogía y de Magisterio se refieren al género, a la opción cursada en COU y a sus calificaciones, tanto previas a su acceso a la Universidad como de 1º de carrera.

Aunque en ambas titulaciones la mayoría de los alumnos matriculados proceden de familias en las que el padre y/o la madre tienen estudios primarios, el porcentaje es ligeramente superior para la Licenciatura en Pedagogía; y para ambas titulaciones es muy bajo el porcentaje de licenciados entre los padres de los estudiantes. Asimismo, es muy elevada la proporción de alumnos cuyas madres no realizan ningún trabajo remunerado, en ambas carreras.

La proporción de mujeres matriculadas es mayor que el de hombres en ambas carreras, así como la de alumnos procedentes de la opción C en COU, pero para ambas variables, las diferencias se acentúan en el caso de Pedagogía. Y por último, señalar que las notas medias de los alumnos, tanto en los cursos de BUP como en el COU, son más altas para Magisterio que para Pedagogía; y lo mismo ocurre con los resultados obtenidos en 1º de carrera: son mejores en los alumnos de Magisterio que en los de Pedagogía.

3.2. Modelización estocástica de la elección de carrera universitaria

A continuación, trataremos de determinar la importancia explicativa que las características académicas y los factores socioeconómicos citados tienen sobre la elección de carrera de los alumnos de las titulaciones analizadas. Para ello, estimaremos un modelo logit binomial en el que la variable dependiente es una variable binaria que refleja la probabilidad de que el individuo correspondiente decida cursar la Licenciatura en Pedagogía o alguno de las seis titulaciones de Magisterio. Así,

CARRERA = 1 si elige estudiar Pedagogía.
0 si estudia Magisterio.

Las variables explicativas serán inicialmente todas las descritas en el apartado anterior, aunque luego iremos eliminando las que no sean significativas en la regresión:

GÉNERO = 1 si es mujer
0 si es hombre
NOTA1 = nota media de las calificaciones obtenidas en 1º de BUP
NOTA2 = nota media de las calificaciones obtenidas en 2º de BUP
NOTA3 = nota media de las calificaciones obtenidas en 3º de BUP
NOTAC = nota media de las calificaciones obtenidas en COU
OPCION = opción cursada en COU, que desagregaremos en otras tres variables dummy para poder incluirla en nuestro análisis:

OPCA/B = 1 si cursó la opción A o la B (consideradas en una sola categoría ya que sus contenidos son muy similares)
0 en otro caso

OPCC = 1 si cursó la opción C
0 en otro caso

EST_P = nivel de estudios del padre, variable de carácter discreto que toma estos valores:
1.- Sin estudios.
2.- Estudios primarios.
3.- Bachiller Elemental o asimilados.
4.- Bachiller Superior o asimilados.
5.- Diplomatura Universitaria o asimilados.
6.- Licenciatura, Ingeniería, Arquitectura o asimilados.

EST_M = nivel de estudios de la madre, variable de carácter discreto que toma los mismos valores que la anterior.

TRAB_P = trabajo del padre, variable de carácter discreto que toma los siguientes valores:
1.- Directores o gerentes de empresas de la Administración Pública.
2.- Técnicos o profesionales asociados a titulaciones universitarias o no, de carácter postsecundario.
3.- Empleados administrativos y trabajadores de servicios.
4.- Trabajadores cualificados en agricultura y pesca.
5.- Trabajadores cualificados en la industria, construcción y minería.
6.- Trabajadores no cualificados.
7.- Profesionales de las Fuerzas Armadas.
8.- Personas que no han tenido trabajo remunerado.
0.- Otros

TRAB_M = trabajo de la madre, variable de carácter discreto que toma los mismos valores que la anterior.

La expresión del modelo que vamos a estimar es:

$$P\,(CARRERA=1) = \frac{\exp\left(\sum_i \beta_i x_i\right)}{1+\exp\left(\sum_i \beta_i x_i\right)}$$

donde x_i son las variables explicativas que definen a cada individuo y β_i los parámetros correspondientes.

La tabla 10 muestra las estimaciones por máxima verosimilitud del modelo logit, realizada con la totalidad de variables explicativas definidas:

TABLA 10
MODELO 1

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	18,5692	2,0623	0,0000
GÉNERO	1,0536	0,2929	0,0003
NOTA1	-0,7206	0,2483	0,0037
NOTA2	-0,6479	0,3007	0,0312
NOTA3	-1,1613	0,3052	0,0001
NOTAC	-0,4801	0,2302	0,0371
OPCA/B	1,1213	0,3529	0,0015
OPCC	0,8902	0,3215	0,0056
EST_P	-0,0234	0,1101	0,8328
EST_M	-0,2313	0,1438	0,1076
TRAB_P	0,0426	0,0546	0,4351
TRAB_M	-0,0041	0,0514	0,9363
Log likelihood	-230,2845		
R^2 conteo	80,08%		

Si eliminamos, paso a paso, en el modelo estimado, las variables con nivel de significatividad superior o igual al 5%, obtenemos finalmente los resultados que se muestran en el tabla 11.

TABLA 11
MODELO 2

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	18,6463	1,9906	0,0000
GÉNERO	1,0655	0,2920	0,0003
NOTA1	-0,7248	0,2468	0,0033
NOTA2	-0,6326	0,2989	0,0343
NOTA3	-1,1863	0,3032	0,0001
NOTAC	-0,4615	0,2296	0,0444
OPCA/B	1,1250	0,3518	0,0014
OPCC	0,9012	0,3190	0,0047
EST_M	-0,2536	0,1053	0,0161
Log likelihood	-231,0375		
R^2 conteo	79,92%		

Al observar la matriz de correlaciones de las variables relevantes en nuestro modelo 2 (tabla 12), vemos que hay una alta presencia de multicolinealidad, sobre todo en las variables relativas a la calificación obtenida por el alumno en los cursos de BUP y COU, y en las opciones cursadas en COU. Esto hace que la varianza de los parámetros estimados sea mayor y, por lo tanto, las estimaciones sean imprecisas.

TABLA 12
MATRIZ DE CORRELACIONES

	NOTA1	NOTA2	NOTA3	NOTAC	OPCA/B	OPCC	EST_M
NOTA1	1,000						
NOTA2	0,748**	1,000					
NOTA3	0,618**	0,705**	1,000				
NOTAC	0,487**	0,554**	0,603**	1,000			
OPCA/B	0,207**	0,236**	0,138**	-0,096*	1,000		
OPCC	-0,125**	-0,193**	-0,148**	0,024	-0,691**	1,000	
EST_M	0,107*	0,057	0,069	0,024	-0,042	0,070	1,000

** La correlación es significativa al nivel 0,01.
* La correlación es significativa al nivel 0,05.

Para intentar reducir este problema, vamos a estimar otros tres modelos alternativos donde no se incluyan las variables que más correlacionan entre sí:

TABLA 13
MODELO 3

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	10,3221	1,2168	0,0000
GÉNERO	0,8121	0,2715	0,0028
NOTA1	-1,5738	0,1773	0,0000
EST_M	-0,2187	0,0977	0,0253
Log likelihood	-267,2465		
R^2 conteo	71,97%		

* En esta regresión ninguna de las variables relativas a la opción cursada en COU aparecen como significativas al 5%.

TABLA 14
MODELO 4

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	12,6195	1,4337	0,0000
GÉNERO	0,8911	0,2773	0,0013
NOTA2	-1,9670	0,2179	0,0000
OPCA/B	0,5054	0,2373	0,0332
EST_M	-0,2136	0,0970	0,0158
Log likelihood	-259,180		
R^2 conteo	74,27%		

TABLA 15
MODELO 5

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	14,5576	1,6152	0,0000
GÉNERO	1,1511	0,2801	0,0000
NOTA3	-2,2482	0,2435	0,0000
EST_M	-0,2445	0,0971	0,0118
Log likelihood	-254,522		
R^2 conteo	74,48%		

* En esta regresión ninguna de las variables relativas a la opción cursada en COU aparecen como significativas al 5%.

TABLA 16
MODELO 6

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	9,9990	1,6152	0,0000
GÉNERO	0,8592	0,2698	0,0014
NOTAC	-1,6342	0,1979	0,0000
OPCC	0,6030	0,2138	0,0048
EST_M	-0,2883	0,0967	0,0029
Log likelihood	-266,366		
R^2 conteo	76,57%		

Aunque vemos que en las estimaciones de los modelos 3 y 5 la variable relativa a la opción cursada en COU no es significativa al 5%, en los modelos 4 y 6 sí son todas las variables significativas. Teniendo en cuenta estos resultados, podemos concluir que en la probabilidad de estudiar una carrera de ciclo largo dentro de las correspondientes a Ciencias de la Educación (Licenciatura en Pedagogía) no ejerce ninguna influencia la ocupación laboral del padre o de la madre, ni los estudios del padre. En cambio, sí tienen influencia otras características del individuo, como su género, la opción cursada en COU y las notas obtenidas antes de entrar a la universidad.

En concreto, el expediente académico del alumno (que entra en el modelo a través de las variables NOTA1, NOTA2, NOTA3 o NOTAC) ejerce una gran influencia sobre la elección de alternativa. Su coeficiente es en todos los casos negativo, lo cual indicará que, a mayor nota media obtenida en 1º de BUP, 2º de BUP, 3º de BUP o COU, el individuo elegirá con menor probabilidad cursar Pedagogía que cualquiera de los títulos de Magisterio. Este hecho es compatible con las conclusiones extraídas del análisis descriptivo de la muestra: la nota media de los alumnos de primer curso es superior en el grupo de Magisterio que en el de Pedagogía.

En segundo lugar, la variable GÉNERO es también altamente significativa en la regresión y tiene un coeficiente positivo, lo que significa que, si el estudiante es una mujer, elegirá la carrera de ciclo largo más probablemente que si es hombre. Esto se corresponde con lo que observábamos en el análisis descriptivo de la muestra: el porcentaje de mujeres en Pedagogía es superior al de Magisterio.

La variable relativa a la opción cursada en COU también influye en la decisión modelizada, aunque de una forma menos decisiva, ya que en los modelos 3 y 5 no es significativa al 5%. En cualquier caso, vemos que los alumnos procedentes de la opción A o B y de la opción C tienen más probabilidad de cursar Pedagogía que los que estudiaron la opción D.

Por último, hemos hallado que la variable EST_M relativa al nivel de estudios de la madre del alumno es también significativa en nuestra regresión; y tiene coeficiente negativo, lo cual indicaría que a mayor nivel de estudios de la madre (tal y como la hemos codificado), menos probabilidad de que el individuo decida estudiar Pedagogía. Con el fin de aclarar más este punto, hemos estimado el modelo, pero incluyendo ahora la variable EST_M no como una variable categórica sino como seis variables binarias: ESTM1, ESTM2, ... ESTM6 que toman valor 1 si EST_M = 1, 2,...6 respectivamente, y 0 en otro caso. Si eliminamos las variables no significativas a un nivel mínimo de significatividad del 5%, obtenemos el modelo 7:

TABLA 17
MODELO 7

Variable explicativa	Coeficiente	Standard Error	Significatividad
C	11,6159	1,4168	0,0000
GÉNERO	0,8665	0,2788	0,0019
NOTA2	-1,9738	0,2191	0,0000
OPCA/B	0,5110	0,2382	0,0320
ESTM1	0,7182	0,3236	0,0264
ESTM2	0,7076	0,2597	0,0064
Log likelihood	-258,045		
R^2 conteo	73,85%		

Es decir, un alumno cuya madre no tiene estudios o tiene un nivel de «Estudios Primarios», tienen mayor probabilidad de estudiar Pedagogía frente a las seis titulaciones de Magisterio ofertadas.

3.3. Modelización de los resultados obtenidos en primer curso

Una vez que sabemos de qué variables depende la elección de una carrera de Ciencias de la Educación de ciclo largo frente a las de ciclo corto, trataremos de analizar qué factores influyen en los resultados académicos obtenidos por los alumnos de ambas titulaciones (Pedagogía y Magisterios) en su primer curso de carrera. A partir de la muestra de que disponemos, estimaremos un modelo logit binomial para cada titulación donde la variable dependiente sea una variable binaria que refleja si el individuo correspondiente ha tenido o no éxito en el primer año de estudios universitarios (definiendo el «éxito» como obtener una nota media, en Junio o Septiembre, de aprobado o superior. Así:

ÉXITO = 1 si «Resultados» $\geq$ 1.
0 si «Resultados» < 1.

Las variables explicativas de las que partimos son todas las descritas anteriormente, aunque posteriormente se irán eliminando aquellas que no resulten significativas.

En la tabla 18 podemos observar las estimaciones por máxima verosimilitud del modelo descrito, efectuada con la totalidad de variables explicativas definidas:

TABLA 18
MODELO 8

Variable	Pedagogía		Magisterio	
	Coeficiente	Significatividad	Coeficiente	Significatividad
C	-13,8974	0,0003	-5,0275	0,0715
GÉNERO	-0,5004	0,2741	-0,2100	0,6601
NOTA1	-0,0217	0,9506	-0,398	0,9226
NOTA2	0,0113	0,9794	-0,5134	0,3263
NOTA3	0,7225	0,0711	1,3146	0,0215
NOTAC	1,7429	0,0006	0,2644	0,2078
OPCA/B	1,1168	0,0188	0,0337	0,9542
OPCC	0,2176	0,6017	0,2243	0,6846
EST_P	0,0141	0,9209	-0,4342	0,0145
EST_M	-0,2429	0,2211	0,4159	0,0648
TRAB_P	-0,0182	0,7841	-0,0549	0,5940
TRAB_M	0,0094	0,8970	0,0736	0,3546
Log likelihood	139,208		84,7285	
R^2 conteo	67,23%		86,97%	

Si ahora eliminamos para el modelo estimado las variables que no son significativas a un nivel de significatividad de al menos el 5%, obtenemos finalmente los resultados que se muestran en la tabla 19:

TABLA 19
MODELO 9

Variable explicativa	Pedagogía	
	Coeficiente	Significatividad
C	-6,1288	0,0050
NOTAC	1,0281	0,0047
OPCA/B	1,0010	0,0006
Log likelihood	164,8485	
R^2 conteo	63,95%	
Variable explicativa	**Magisterio**	
	Coeficiente	Significatividad
C	-6,1035	0,0085
NOTA3	1,1454	0,0009
Log likelihood	97,158	
R^2 conteo	84,77%	

A partir de estos últimos modelos estimados, vemos que el género y el nivel de estudios y trabajo de los padres son variables que no ejercen ninguna influencia sobre la probabilidad de éxito de un alumno que inicia sus estudios en la Facultad de Ciencias de la Educación.

En concreto, para la Licenciatura en Pedagogía la probabilidad de éxito de un alumno de 1º va a depender de si ha cursado la opción A o B en COU, de la nota media obtenida en dicho curso (los coeficientes de las variables OPCA/B y NOTAC han resultado ser positivos). En el caso de Magisterio, sólo es significativa la variable NOTA3, que como también observamos aparecen con un coeficiente positivo, indicando que a mayor nota media en 3º de BUP, mayor probabilidad de éxito.

El hecho de que las variables que reflejan las calificaciones medias obtenidas en los demás cursos de Enseñanza Secundaria no aparezcan como significativas no significa que no contribuyan a determinar la probabilidad de éxito o fracaso, sino que su influencia se manifiesta posiblemente a través de la variable presente, con la cual existe una elevada y significativa correlación (véase tabla 12).

4. CONCLUSIONES

Los alumnos de Pedagogía presentan el perfil siguiente:

El 52,3% sus padres han cursado estudios primarios y sus madres el 61,4%; el 21,8% de los padres están empleados en trabajos administrativos y de servicios y el 71,5% de las madres no ha tenido trabajo remunerado; el 52,3% ha cursado la opción C de COU; el 85,3% de la muestra son mujeres. En cuanto a las notas de BUP y COU, la media supera en cada uno de los tres cursos 6 puntos. La nota de primero de carrera se ha distribuido de la siguiente forma: un 48,8% de los alumnos han aprobado el curso, el 41,5% lo han suspendido y el resto se distribuye entre notable y sobresaliente.

El perfil de los alumnos de Magisterio es el siguiente:

El 42,7% de los padres y el 48,5% de la madres sólo han cursado estudios primarios; el 25,5% de los padres está empleados en puestos de la administración y los servicios y el 73,6% de la madres nunca ha ocupado un puesto de trabajo remunerado. En cuanto a la opción estudiada en COU los porcentajes entre las cuatro opciones se encuentra mucho más equidistribuida apareciendo el porcentaje más alto también en la opción C, con un 38,5%. También en Magisterio el porcentaje de mujeres que estudian estas titulaciones es mayor que el de hombres, con un 74,6%. La nota media en los tres cursos de BUP y COU se aproxima al 7, siendo la más alta la de primero que supera este valor. En cuanto a los resultados en primero de carrera observamos que el 62,1% de los alumnos supera el aprobado, frente al 15,2% que no lo supera y el 22,2% que alcanza una media de notable.

Cada una de esta variables estudiadas ha resultado ser significativa entre un grupo de alumnos y otro, Pedagogía frente a Magisterio, en el género, la opción cursada en COU y las calificaciones obtenidas tanto antes de entrar en la universidad como en primero de carrera.

En la modelización econométrica de la elección de carrera que se ha realizado observamos que las variables que más peso tienen en la elección de la licenciatura o la

diplomatura han sido el género, la opción cursada y la nota media obtenida antes de entrar en la universidad, resultando que a mayor nota antes de entrar en la universidad más probablemente elija la titulación de Pedagogía. El género también tiene su influencia: ser mujer le llevará con mas probabilidad a elegir Pedagogía que Magisterio. La opción de COU también resulta determinante en la elección: el haber cursado la opción C influirá más en elegir ,Pedagogía frente a Magisterio. Los estudios de la madre también parecen haber influido en la elección entre ambas titulaciones de tal forma que a mayor nivel de estudios de la madre menos probabilidad existe de que el alumno elija la carrera de Pedagogía.

Un vez que conocemos qué variables son las que determinan con mayor fuerza la elección de titulación en la Facultad de Educación hemos procedido a analizar qué variables son las que más influyen en los resultados académicos en primero de carrera. En los alumnos de Pedagogía la opción cursada en COU y la nota media obtenida son las variables más influyentes; en cambio en Magisterio la influencia mayor la tiene la nota media obtenida en tercero de BUP.

Aunque el trabajo que presentamos se refiere exclusivamente a las titulaciones de un sólo centro, queremos destacar la importancia que la ampliación del tipo y número de las variables que se han incluido en la Guía del Plan Nacional de Evaluación de la Calidad de las Universidades debería ir más allá de las establecidas en dicho documento. En primer lugar porque antes de empezar a evaluar una titulación, éstas y otras variables previas de los alumnos y del contexto, ayudarían a valorar cual es su posición en el contexto general de la universidad. Es decir, la evolución y resultados de una titulación vendrán condicionados, nos guste o no, por las variables de entrada, unas procedentes del alumno, otras del contexto, otras de las exigencias impuestas desde la propia universidad para acceder a la titulación, etc. Por consiguiente, entendemos que éstas y otras variables deberán seguir siendo analizadas para un mejor proceso en la autoevaluación de la enseñanza.

BIBLIOGRAFÍA

ALBERT VERDÚ, C. (1988). *La Demanda de Educación Superior en España: 1977-1994*. Ministerio de Educación y Cultura. CIDE.

APODACA, P. y LOBATO, C. (Eds.) (1997). *Calidad en la Universidad: Orientación y Evaluación*. Barcelona: Laertes.

ARROW, K.J. (1973). *Higher Education as a Filter*. Journal of Public Economics, julio, vol. 2.

ASHENFELTER, O. y HAM, J. (1979). *Education, Unemployment and Earnings*. Journal of Politcal Economy, 87.

BECKER, G.S. (1964). *Human Capital: A Theorical and Empirical Analysis, with Special Reference to Education*. NationalmBureau of Economic Research, New York.

BLAUG, M. (1976). *The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey*, Journal of Economic Literature, 14.

BRICALL, J.M. (2000). *Informe Universidad 2000*.

CAMPBELL, R. y SIEGEL, B.N. (1967). *The Demand for Higher Education in the United States, 1919-1964*. American Economic Review, 57.

DE MIGUEL, M. (1991). *Utilización de indicadores en la evaluación de la docencia universitaria*. En De Miguel, M.; Mora, J.G. y Rodríguez, S. La evaluación de las instituciones universitarias. Madrid: Consejo de Universidades.

HECKMEN, J. J. (1976). *A Life-cycle Model of Earnings, Learning and Consumption*. Journal of Politcal Economy, 84.

MADDALA, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrices*. Cambridge University Press. Cambridge.

MEC (1998). *Plan Nacional de Evaluación de la Calidad de las Universidades*. Madrid: Consejo de Universidades, Secretaría General.

MEC (1999). *Indicadores en la Universidad: Información y decisiones*. Madrid: Consejo de Universidades. MEC.

MICHAEL, R.T. (1973). *Education in Non-market Production*. Journal of Politcal Economy, 80.

MINGER. J. (1974). *Schooling, Experience and Earnings*. NationalmBureau of Economic Research, New York.

MORA, J.G. (1997). *La demanda de educación superior: una revisión de estudios empíricos*. Revista de Educación, 288.

NICKELL, S. (1979). *Education and Lifetime Patterns of Unemployment*. Journal of Politcal Economy 87.

PSACHAROPOULOS, G. (1980). *Returns to Education: An Updated International Comparison*. Comparative Education 17.

PSACHAROPOULOS, G. (1985). *Returns to Education: A futrher International Update and Implications*. Journal of Human Resources, 20.

RILEY, J.G. (1979). *Testing the Educational Screening Hipothesis*. Journal of Politcal Economy 87.

SCHULTZ, T.W. (1960). *Capital Formation by Education*. Journal of Politcal Economy 68.

SCHULTZ, T.W. (1963). *The Economic Value of Education*. Columbia University Press, New York.

SMITH, A. (1987). *Investigación sobre la naturaleza y causas de la riqueza de las naciones*. Fondo de Cultura Económica, México.

SPENCER, M. (1973). *Job Market Signaling*. Quarterly Journal of Economics, 87.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 119-146

AUTOCONCEPTO, AUTOESTIMA Y RENDIMIENTO ACADÉMICO EN ALUMNOS DE 4º DE E.S.O. IMPLICACIONES PSICOPEDAGÓGICAS EN LA ORIENTACIÓN Y TUTORÍA

Miguel Ángel Broc Cavero*
Ministerio de Educación y Ciencia
Diputación General de Aragón

RESUMEN

Este trabajo muestra las relaciones y el peso relativo que determinadas variables antecedentes del autoconcepto mantienen sobre la autoestima global en alumnos de 4º curso de E.S.O. También se analizan las relaciones que determinadas variables psicológicas y pedagógicas tienen sobre el rendimiento escolar final.

Ciento veinte alumnos formaron parte de esta investigación. Las instrumentos elaborados por S. Harter y utilizados en esta investigación fueron: el «Self-Perception Profile for Adolescents» (1988), que mide variables de autoconcepto y autoestima propias de este nivel evolutivo; el «Social Support Scale for Children» (1985 c); la «Scale of Intrinsic Versus Extrinsic Orientation in the Classroom» (1980); el Inventario de Hábitos de Estudio de F.F. Pozar

* Doctor en Psicología; Diplomado en Ciencias de la Educación. Jefe de Dto. de Orientación del I.E.S. «El Portillo» de Zaragoza.-DGA-MEC, y profesor tutor de la Facultad de Psicología en el Centro de la UNED de Barbastro (Huesca).

Dirección del autor: Miguel Ángel Broc Cavero. C/Flora Tristán, 5, 2º B. 50015. Zaragoza. 976-740049. Dirección del trabajo: I.E.S. «El Portillo». C/Juan XXIII, 3. 50010. Zaragoza. 976-336433. E-mail: mabroc@teleline.es

BEST COPY AVAILABLE

(1983); una medida de la inteligencia (Cattell, 1989), las horas diarias de estudio; y el rendimiento en la 1ª, 2ª y 3ª evaluación en las materias propias de este nivel educativo.

Los análisis efectuados por medio del programa estadístico BMDP muestran unos resultados que suelen coincidir con los propuestos por Harter, y aportan interesantes sugerencias muy útiles para generar estrategias de investigación y de intervención en este campo.

Palabras clave: *autoconcepto, autoestima, dimensiones, apoyo social, hábitos de estudio, inteligencia, rendimiento académico, materias de la E.S.O., intervención.*

ABSTRACT

This work shows the relationships and the relative weight that given variables antecedents of the self-concept maintain on the global self-esteem in pupils of fourth course of Obligatory Secondary Education. Also they are analyzed the relationships that given pedagogic and psychological variables have on the scholastic achievement.

A hundred twenty pupils formed part of this research. The instruments elaborated by S. Harter and used in this work were: the «Self-Perception Profile for Adolescents» (1988), that it measures variables of self-concept and self-esteem; the «Social Support Scale for Children» (1985); the «Scale of Intrinsic versus Extrinsic Orientation in the Classroom» (1980); the «Inventory of Study Habits» of F.F. Pozar (1983); a measure of the intelligence (Cattell, 1989); the daily study hours, and the achievement in the 1ª, 2ª and 3ª evaluation in the own matters of this educational level.

The analysis effected by the statistic program BMDP shown some results that tend to coincide with the proposed by Harter, and provide interesting very useful suggestions to generate investigation strategies and of intervention in this field.

Key words: *self, self-concept, self-esteem, dimensions, social support, study habits, intelligence, matters of E.S.O., academic achievement, education, intervention.*

I. INTRODUCCIÓN CONCEPTUAL

Introducción conceptual al modelo cognitivo-evolutivo del Autoconcepto y la Autoestima de S. Harter

En los últimos años se está produciendo un resurgimiento de los estudios sobre el conocimiento del «self» (sí mismo) que aportan nuevos datos que permiten progresar en la construcción de modelos teóricos.

No nos es posible detenernos ahora en el análisis de otros modelos teóricos igualmente importantes propuestos en torno a los constructos autoconcepto (AC) y autoestima (AE), (Coopersmith, 1967; Damon y Hart, 1988; Lewis y Brooks-Gunn, 1979b; Piers-Harris 1969; Rosenberg, 1979), ni en la distinción entre dichos constructos (Broc, 1991, 1994; García, 1983), como tampoco en los instrumentos de medida más utilizados, sus propiedades psicométricas (Hughes, 1984), y la investigación generada en este campo (Broc, 1998; Harter & Broc, 2000).

Tampoco entraremos en el análisis de la extensa intervención desarrollada en esta área, escasa en nuestro país, aunque con un desarrollo incipiente. En este sentido pueden servir de guía algunas de los trabajos realizadas por Broc (1995, 1996), Broc y García Torres (1993), Machargo (1991), y más relacionado con la motivación académica, el trabajo de González Torres (1997).

Consideramos que el modelo de S. Harter es uno de los más completos y relevantes actualmente en el estudio del autoconocimiento y la autoestima y que puede servir de guía para la generación de estudios empíricos en este campo. Las razones por las que nos adherimos a este modelo, sin minusvalorar la contribución de otros enfoques, son las siguientes: a) Hay muy pocos estudios realizados en nuestro país. b) Su modelo teórico es bastante comprensivo e integrador, incluyendo supuestos recogidos por otros investigadores y modelos previos, y adoptando una perspectiva cognitivo-evolutiva. c) Es operativo, ya que la aplicación de sus escalas es relativamente fácil y rápida. d) Utiliza un material atractivo y de fácil aplicación que se adapta a la edad y nivel educativo de los sujetos. e) Permite la comparación de los perfiles, haciendo posible la detección de los sujetos que, teniendo una baja autoestima, no puntúan igual en las dimensiones del autoconcepto. f) El número de investigaciones llevadas a cabo por esta autora es bastante extensa e integra y recoge otras líneas de estudio realizadas recientemente en este campo. g) Se ha llevado a cabo recientemente la adaptación española de uno de los instrumentos de medida elaborados inicialmente por Harter, para la población adolescente, obteniendo unos índices de fiabilidad y validez muy aceptables. (Harter & Broc, 2000, COSPA, en prensa).

Los modelos de autoconcepto y autoestima en la literatura científica se han distinguido por el número de dimensiones a evaluar, y lo más importante de ello es, si el «Yo» se ha visto, como un constructo unidimensional, o como autoevaluaciones que son más multidimensionales, en su naturaleza. Cada uno de estos modelos ha generado una estrategia de medida relacionada y acorde a los supuestos teóricos que propugna. Aunque se han formulado diversas aproximaciones al estudio del autoconcepto y autoestima (Broc, 1991), creemos que los modelos más interesantes en este campo han sido propuestos por Coopersmith (1967), Piers-Harris (1969), Rosenberg (1979), y el propuesto por Harter (1986, 1987, etc.), modelo que ha servido de base en la elaboración de este instrumento.

La gran mayoría de los modelos propuestos, además de no diferenciar entre los constructos autoconcepto y autoestima (García Torres, 1983), se inclinaban por una medida unidimensional de uno o de los dos constructos. Esta naturaleza unidimensional ha sido puesta en tela de juicio por recientes teóricos que argumentan que tal aproximación oculta importantes dimensiones evaluativas que los sujetos hacen acerca de su aptitud en diferentes dominios de sus vidas, además de la inadecuación del instrumento para la utilización en los diversos periodos evolutivos. El modelo de Susan Harter (1985; 1987) representa una integración de las perspectivas unidimensionales y multidimensionales, subrayando la importancia de juicios globales de estima o autovalía, además de la evaluación de competencias percibidas en dimensiones específicas. De esta forma la AE global se investiga directa e independientemente por un

conjunto de ítems que miden diferentes juicios individuales de valía como persona (similar a la conceptualización de Rosenberg).

Esta aproximación a la autoestima global es, por tanto, totalmente diferente de los procedimientos de aquellos autores que han tenido que definir de forma operativa el autoconcepto general como la suma o media aritmética de las respuestas del niño/a o alumno/a a un amplio conjunto de ítems que miden diversas áreas de contenido, (por ejemplo, la medida de la autoestima elaborada por Coopersmith). Es decir, nosotros no nos adherimos al enfoque de que la autoestima global es mejor investigada sumando respuestas dadas a un conjunto de ítems cuyas preguntas giran en torno a una amplia variedad de auto-descripciones. Más bien, nosotros creemos que los propios sentimientos de autoestima deben ser medidos directamente, preguntando acerca de la autoestima en sí misma considerada. Por lo tanto, queremos que nuestros ítems animen a los adolescentes a pensar acerca de la percepción global de su valía como personas. De esta forma, no tendremos que inferir la autoestima de la suma o media aritmética de sus respuestas a cuestiones específicas relativas a sus destrezas o características determinadas.

Es necesario dejar claro que la **autoestima global** no es una medida de la competencia general. Algunos otros investigadores que han empleado el instrumento original han malinterpretado estas subescalas, considerándolas como un índice más amplio de competencia. No obstante, si se investiga la autoestima global separada o independientemente de las dimensiones de competencia específicas, se puede entonces examinar las relaciones entre la autoestima global y las competencias percibidas en las dimensiones específicas. De hecho, ha sido necesario determinar tanto los antecedentes como los correlatos de la autoestima global (Harter, 1986, 1987, 1989). Como cita, esta autora ha encontrado evidencia de los dos determinantes, basándose en las contribuciones teóricas de dos grandes teóricos del «self» («yo») como son James (1892) y Cooley (1909).

Para W. James, la autoestima global representaba la proporción de éxitos personales partido por sus pretensiones o aspiraciones hacia el éxito en las diferentes dimensiones de su vida. Harter ha operacionalizado esta formulación y ha llegado a afirmar que la competencia o adecuación en dimensiones consideradas importantes para la persona, es fuertemente predictivo de la autovalía o autoestima. Un segundo predictor de la autoestima, igualmente potente, es la visión positiva que los «otros significativos» mantienen hacia el «yo». La principal contribución de Cooley fue que el «yo» representa las valoraciones reflejadas de los otros significativos. Los hallazgos de Harter indican claramente que la visión percibida de los «otros» ejerce un impacto directo sobre la visión global que uno tiene de sí mismo. No obstante, esta fuente de autoestima es independiente de la anterior, relativa a la competencia percibida en dimensiones consideradas importantes. Una discusión completa de los efectos, incluyendo un modelo causal de los determinantes y correlatos de la autoestima, pueden encontrarse en Harter (1986).

La versión de este instrumento en muestras americanas presenta unos índices de fiabilidad que oscilan entre los valores de .74 y .91 en diferentes muestras utilizadas, y arroja una estructura de factor con ocho subescalas separadas que miden dimensiones

específicas y otra dimensión separada de autoestima global. Las subescalas son la competencia académica, aceptación social, competencia atlética, apariencia física externa, comportamiento, competencia en el trabajo, Amigos Íntimos, Atracción Romántica y Autoestima Global. Esta última presenta correlaciones que oscilan entre .31 y .73 con las ocho dimensiones anteriores. Existe una versión adaptada a la población española llevada a cabo por Broc (Harter & Broc, 2000, en prensa). Los coeficientes de fiabilidad oscilan entre .71 y .92 en las diferentes muestras, y también se encuentra una estructura de 8 factores con altas correlaciones dentro de cada uno de ellos, y bajas correlaciones cruzadas.

La definición de autoconcepto que planteamos en este trabajo es la de «una representación interna y cognitiva que el sujeto tiene en una serie de dimensiones que han sido halladas principalmente mediante técnicas de análisis factorial y análisis de cluster». Por otra parte, en esta exposición adoptaremos el término autoestima o autovalía indistintamente, y nos identificaremos con la definición que del término hace Harter (1983, 1987). Por autoestima global entendemos los «diferentes valores que uno coloca sobre el «Yo» como persona, un juicio que puede ser contrastado con otras evaluaciones o juicios sobre la competencia de uno mismo en áreas o dominios específicos. La autoestima sería, por decirlo así, el componente afectivo-evaluativo del autoconcepto (García Torres, 1983). Estos elementos indagan hasta qué punto el individuo se gusta a sí mismo como persona, es feliz con la forma en que dirige su propia vida, etc. Por otra parte, subescalas separadas miden evaluaciones de áreas o dimensiones específicas que proporcionan un perfil multidimensional, como puede verse en las escalas de medida elaboradas por esta autora en las que se incluyen diferentes dimensiones que varían en diferentes periodos evolutivos, comenzando en la infancia media y extendiéndose hasta la edad adulta. Separando conceptual y empiricamente los juicios de áreas o campos específicos sobre la competencia o aptitud del juicio más global de la valía individual como persona, Harter ha sido capaz de determinar las relaciones que las competencias específicas aportan a la autoestima global, es decir, ha podido especificar la forma por la cual las dimensiones son sopesadas y combinadas para producir un determinado sentido de autoestima global. Se ha puesto especial énfasis en que el AC puede ser conceptualizado a lo largo de un «continuum» de juicios muy específicos a muy globales acerca del «Yo», y que éstos no son necesariamente excluyentes.

La aproximación de Harter es muy similar a la de Rosenberg. Harter aboga por un sistema conceptual en el cual las dimensiones de la autoevaluación son de algún modo organizadas de forma jerárquica y tal vez, sopesadas de forma diferencial. Además, la aproximación evolutiva de Harter no implica que deba inferirse la AE global de alguna combinación de autoevaluaciones medibles en el segundo o tercer nivel, aunque se haya hecho así en muchos otros estudios.

En un trabajo realizado por Harter y Pike (1984), se demostró que los niños de 4 a 7 años pueden hacer juicios fiables acerca de los cuatro dominios siguientes: competencia cognitiva, competencia física, aceptación de los compañeros y aceptación de los padres. Respecto a la estructura del AC/AE en la *infancia media*, Harter afirma que en esta edad, no solamente emerge una mayor amplitud de áreas que se van diferenciando, sino la habilidad para hacer juicios de la valía de uno mismo como persona. Los

niños en edades comprendidas entre los 8 y los 12 años diferencian claramente cinco dominios, que a su vez, están contenidos en dicha escala: competencia escolar, competencia atlética, aceptación social de los compañeros y apariencia física, incluyéndose una escala de autovalía global independiente de las autoevaluaciones.

Respecto a la estructura del AC/AE en *adolescentes* puede observarse una mayor articulación y diferenciación. La prueba inicial utilizada por Harter, «Self-Perception Profile for Adolescents» (1988), que utilizamos en este trabajo, incluye tres nuevas dimensiones: amigos íntimos, atracción romántica y competencia en el trabajo, que sumadas a las cinco señaladas anteriormente hacen un total de ocho. A estas se añade una subescala de autoestima global.

En la estructura del AC/AE en sujetos de universidad todavía se observa una mayor proliferación de dimensiones tanto más articuladas como diferenciadas. (Neeman y Harter (1986), así como en la población adulta (Messer y Harter, 1989).

La estructura de cada una de estas subescalas refleja diferencias evolutivas a lo largo del ciclo vital, que se han conceptualizado como intereses cambiantes e influencias del contexto así como diferencias que incluyen también procesos cognitivo-evolutivos.

Un modelo de los determinantes de la autoestima. La operacionalización de la teoría de W. James y Cooley

Habiendo conceptualizado el AC como un conjunto de juicios sobre dominios específicos acerca de la competencia y adecuación en relación a uno mismo, además del juicio global de la valía como persona, es interesante examinar las relaciones entre estos dos tipos de comentarios autoevaluativos. Por ejemplo, ¿la autoestima global representa meramente una combinación sumativa de juicios de dominio específico, o es algo más amplio en su conjunto, o diferente, de la suma de sus partes?

Más que reinventar algunos supuestos conceptuales, Harter vuelve sobre dos grandes teóricos clásicos en el AC, como son W. James (1892) y Cooley (1902), para pasar posteriormente a estudiar cómo es posible operativizar todo lo posible dichos supuestos de cara a dar a la investigación todo el carácter científico y experimental que se merece.

Para probar **el modelo de James** con niños (a partir de 8 años) era necesario operacionalizar tanto la *competencia o éxito* en una determinada dimensión, como la *importancia concedida* por el sujeto a la misma, (el concepto de «pretensión» de James), a lo largo de los diferentes dominios de la vida del niño.

La competencia percibida o éxito lo ha definido sobre la base de la competencia percibida o adecuación obtenidas en los instrumentos de medida correspondientes en los distintos niveles evolutivos. El constructo de «pretensiones» lo ha operacionalizado como la importancia del éxito en cada uno de los dominios establecidos previamente. Por lo tanto su fórmula éxito/pretensiones permite obtener juicios de competencia/adecuación y clasificaciones en importancia que se pueden comparar en cada dominio respectivamente, en un intento de comparar si son las relaciones entre competencia e importancia del éxito las que predicen la autoestima o autovalía global del individuo, que es investigada por su propio conjunto de ítems. (Harter, 1985a).

En relación a obtener la proporción de que hablaba James entre éxitos (competencia percibida) y pretensiones (importancia del éxito) se calculó una puntuación de *discrepancia total* (competencia menos importancia) promediada a través de, justamente aquellos dominios que el sujeto consideraba importantes. (En la formulación de James, la competencia alta o baja, en áreas no consideradas importantes no debería influir en la autoestima global del individuo). Un bajo nivel de AE vendría determinado por una puntuación en discrepancia relativamente alta, es decir, que las puntuaciones en importancia del éxito en ese dominio excederían los juicios de competencia/adecuación.

Por otra parte, los individuos con alta AE obtendrían una puntuación próxima a cero, por lo que las puntuaciones obtenidas en la jerarquía de competencias percibidas serían congruentes con la jerarquía de los juicios concernientes a la importancia del éxito en dichos dominios.

En relación a la **operacionalización de la teoría de Cooley**, Harter define el constructo de las «opiniones de los otros hacia el Yo» como el grado por el cual el individuo siente que los otros conocen su valía como persona. De esta forma se puede medir la percepción positiva percibida así como el apoyo emocional que los demás exhiben hacia el «Yo» con el objeto de determinar si esto predice la visión o estima que un individuo mantiene hacia sí mismo, correspondiente a la definición de autoestima. Se han identificado cuatro fuentes de apoyo social: padres, profesores, compañeros de clase y amigos íntimos.

Estos ítems miden hasta qué punto el individuo percibe o siente que estos «otros significativos» tratan al sujeto como una persona valiosa, sienten que es importante, escuchan lo que él/ella dice o hace, les gusta la forma en que la persona se comporta, tienen cuidado de sus sentimientos, etc. (Harter, 1985 c).

Es preciso destacar que el soporte o apoyo social puede ser, y ha sido, definido de diferentes formas. El intento particular de esta autora fue operacionalizarlo de forma que la visión de los otros significativos fuera análoga a la visión que uno mantiene de sí mismo, es decir, hacer estos dos constructos tan conceptualmente paralelos como fuera posible. Esto es, si uno desea probar la hipótesis de que la autoestima incluye la incorporación de las actitudes de los otros significativos hacia el «Yo», el contenido de ambas medidas debería ser teoricamente similar.

Existe una relativamente amplia evidencia empírica que apoya este modelo en los distintos periodos evolutivos, y parece que un modelo sumativo es el que parece detectar los efectos en los que éstos constructos se combinan para producir el nivel de autoestima individual. (Harter, 1987, 1988).

Los hallazgos revelan que ambos constructos («competencia percibida en áreas importantes» y apoyo social de los «otros significativos»), son críticos en la determinación de la autoestima y que uno no compensa al otro, y no hay interacción. Esto hace que un individuo pueda sufrir una baja autoestima si hay congruencia en la puntuación competencia/importancia pero sin embargo puntúa bajo en las escalas de apoyo social, y a la inversa, no hay congruencia en la primera aunque goce de un saludable apoyo social. Por lo tanto, el sujeto que goza de una buena y alta AE es quien posee una combinación de alto apoyo social y baja puntuación en discrepancia. El patrón opuesto sería característico del niño con más baja autoestima.

Los determinantes de la autoestima en la adolescencia

De acuerdo con la noción de James de que la discrepancia entre la competencia de un sujeto en una dimensión y la importancia que le concede es un determinante de la autoestima, Harter, (1985) operacionalizó la discrepancia como el promedio de las discrepancias en los dominios de cada periodo evolutivo. Al igual que se ha realizado en esta investigación, Harter quiso saber si ciertos dominios contribuían más que otros como predictores de la AE, por lo que examinó las correlaciones de las puntuaciones discrepantes, calculadas separadamente para cada dimensión, y la puntuación en AE, y esto lo hizo en muestras comprendidas en los distintos niveles evolutivos.

Los hallazgos revelan que ciertas áreas contribuyen sistematicamente más a la autoestima que otras. Centrándonos en la etapa que estamos considerando (16-17 años), Harter (1988), halló que las correlaciones entre algunas de las diversas dimensiones con la autoestima variaban significativamente (r = .57 apariencia física externa); (r = .45 aceptación de los compañeros); (r = .36 competencia escolar); (r = .24 competencia atlética) y conducta (r = .26), encontrándose resultados similares en muestras españolas (Harter & Broc, 2000).

El impacto del mayor peso de la apariencia física sobre la AE es explicado por la influencia de los medios de comunicación de masas en los que se exalta la imagen corporal y el atractivo físico. Como Elkind (1979) ha apuntado recientemente, la importancia de la apariencia física entre la gente joven ha sufrido un aumento creciente en los últimos años. Este punto será tratado posteriormente en esta investigación.

La correlación hallada entre la apriencia física y la AE en el mundo adulto y en los niños superdotados son relativamente altas (r = -.65 y r = -.67) respectivamente. Esta consistencia evolutiva ha hecho pensar que la apariencia física puede representar la manifestación de un «Yo exterior», mientras que la autoestima representaría el «Yo interior» ambos definidos globalmente.

Desde una perspectiva evolutiva, la primacía del «Yo físico» ha sido igualmente documentada en aquellos estudios que versan sobre el surgimiento del «yo» en el niño. (Lewis y Brooks-Gunn, 1979 b). Por otra parte, algunos estudios (Maccoby y Martin, 1983) revelan que los «otros significativos» reaccionan al niño y al bebé sobre la base de su apariencia física y/o atractivo físico.

De esta forma, desde una edad muy temprana el «yo físico externo» podría ser una dimensión saliente que no necesariamente estaría unida a las capacidades individuales, y que provoca reacciones psicológicas de los demás que pueden ser perfectamente encuadrables en el sentimiento emergente del «Yo interno» de cada individuo. Este análisis especulativo implicaría que el «Yo externo» podría desarrollarse previamente, y además, influiría en el «Yo psicológico interno» aunque es posible que la dirección de las relaciones entre ambos cambien con el desarrollo. Es una cuestión que debe investigarse tanto evolutiva como diferencialmente el hecho de que en ciertos periodos evolutivos pueda ser posible que el impacto de uno sobre el otro difiera a través de los individuos.

El análisis de la contribución relativa de las fuentes de apoyo social en sujetos de 8 a 16 años, en el estudio de Harter (1987, p. 230), indica que el patrón es relativamente

similar. El apoyo social de los padres y de los compañeros de clase dieron la contribución más alta con respecto a la AE (r entre .42 a .45). A estas siguieron las puntuaciones correspondientes al apoyo social de los amigos íntimos y el de los profesores («r» entre .30 a .27), mostrando que estas fuentes más posteriores tienen menos impacto sobre la AE. Resultados en muestras españolas pueden consultarse en Harter y Broc, 2000 (en prensa).

En base a estos datos se sugiere que, aunque los compañeros tienen un papel más amplio en este periodo evolutivo, los padres son todavía elementos críticos en su impacto respecto al sentimiento de valía de sus hijos.

Por otra parte, el hecho de que los compañeros de clase sean una mayor fuente de apoyo social que un amigo íntimo estriba en que el reconocimiento de los compañeros en un mayor número de dimensiones públicas o sociales parece más crítico que la opinión personal y única de un amigo íntimo. Más bien, se tiende hacia fuentes más objetivas de «feedback» que concedan validez al sentido del «Yo». No obstante, los procesos subyacentes a la influencia de los compañeros de clase y los amigos íntimos requiere de mayor investigación, ya que este patrón no se limita exclusivamente al periodo de la infancia y adolescencia, ni quizás, a todas las culturas. Un patrón similar se observa cuando trasladamos este análisis a sujetos de universidad (Harter, 1987).

De especial interés es el hecho de que las dos dimensiones más predictivas de la AE global no reflejan competencias per se, en el sentido de destrezas o habilidades, a pesar de que en esta población la competencia es una dimensión indudablemente cierta. De ello se deduce, según Harter, que el «Yo exterior o social» del individuo reflejado por la apariencia física y la aceptación de uno mismo por los compañeros, puede jugar un papel funcional más importante que el «Yo social interno o privado».

Relaciones entre el autoconcepto y el rendimiento académico

En lo que respecta a las relaciones entre el autoconcepto y el rendimiento académico, Harter (1983) también ha planteado en su modelo algunos mecanismos que intentan explicarlo. Los hallazgos revelan relaciones positivas entre la competencia académica percibida y el nivel de logro escolar conseguido. Los alumnos que perciben alta su propia competencia escolar tienen más probabilidad de alcanzar el éxito académico, mientras que aquellos cuya competencia percibida es más bien baja tienen más probabilidad de ser peores estudiantes.

En los últimos años, ha habido un intento por parte de los teóricos, de examinar un sistema más amplio de relaciones entre estos constructos. Dentro del campo académico, por ejemplo, los investigadores se han centrado principalmente en las reacciones *afectivas o emocionales* que los alumnos tienen en respuesta a su propia actuación escolar o académica, así como también en la *motivación* de los mismos para el aprendizaje escolar. Los estudios realizados sugieren un sistema de «reacciones en cadena» en forma de espiral que se retroalimenta positiva o negativamente. (Harter, 1988 a).

Recientemente han surgido una serie de modelos teóricos que intentan explicar las relaciones entre autoconcepto académico y rendimiento escolar y que incluyen componentes cognitivos —expectativas, creencias, metas y valores— como elementos cla-

ves para explicar el proceso motivacional. (McCombs, 1992; Weiner, 1990). La Teoría Atribucional de la Motivación de Weiner (1985) postula que la necesidad de rendimiento está mediada por las atribuciones las cuales influyen en las expectativas de éxito, en las percepciones de control y competencia, en las reacciones afectivas y en la conducta motivacional. Otros autores como Dweck (1986) y Nicholls (1984a) se centran en el papel de las metas y postulan que la motivación puede ser adaptativa o desadaptativa según las orientaciones de meta de los estudiantes. Por otra parte, Bandura (1977) subraya el papel de las percepciones de autoeficacia y su influencia en el nivel de motivación. Covington (1984), en su teoría de la Autovalía afirma que las bajas percepciones de competencia inducen a los estudiantes a orientar su conducta más hacia la defensa del yo que hacia un adecuado aprendizaje. Finalmente, Pintrich y De Groot (1990 a) señalan que las principales teorías actuales de la motivación académica se articulan en torno a componentes de expectativas, componentes de valor y componentes afectivos.

Todas estas líneas apuntan hacia la preponderancia de los factores de metacognitivos de autorregulación en el aprendizaje, junto a la integración de aspectos motivacionales que guían y orientan la conducta académica. No es el objetivo de este trabajo analizarlos en este lugar aunque existen publicaciones muy interesantes que el lector puede encontrar, entre otros, en Connell, (1990), Csikszentmihalya (1990), Deci y Ryan (1992), DeCharms (1994), Eccles & Midgley, (1989); Paris y Turner (1994), Schunk (1994), Urdan & Maher (1995), Zimmerman (1994), y en nuestro país Alonso Tapia (1992a, 1992b, 1995), González Torres y Tourón, (1992), Núñez y González (1994).

II. OBJETIVOS DE LA INVESTIGACIÓN

Después de esta introducción estamos en condiciones de plantearnos las siguientes cuestiones: ¿Cuál es el peso relativo que las diferentes dimensiones del autoconcepto tienen sobre la autoestima global del alumno al final de la E.S.O.?, ¿se observa un patrón similar o diferente al encontrado por Harter en la población americana?, ¿qué influencia tiene el apoyo social y la discrepancia sobre la autoestima?, ¿cuáles son los mejores predictores de la autoestima a los 16-17 años, de cara a poder intervenir a tiempo sobre los mismos?, ¿cuáles son las variables que mejor predicen el rendimiento académico en esta etapa evolutiva?, ¿Tienen relación las variables de AC y AE sobre el rendimiento, o son únicamente variables pedagógicas las que tienen más peso?, ¿hacia dónde podría dirigirse la investigación y la intervención en este campo?

Esta investigación se dirigió hacia los siguientes objetivos: a) conocer el peso relativo que determinadas dimensiones antecedentes del autoconcepto tienen sobre la autoestima global del alumno de final de la ESO, es decir, vislumbrar qué variables predicen mejor una alta o baja autoestima en esta edad. b) Comprobar si los resultados obtenidos son consistentes con el modelo teórico propuesto por Harter, y si dichos datos confirman o refutan la teoría. c) Examinar las relaciones entre variables de autoconcepto y variables académicas sobre el rendimiento final del alumno, es decir, buscar las variables independientes que mejor predicen el rendimiento final del alum-

no en el último curso de la ESO. d) En base a los resultados obtenidos plantear estrategias de intervención y de investigación en este campo.

III. DISEÑO Y MÉTODO

Sujetos

Ciento veinte alumnos/as de 4° de la E.S.O. (16-17 años) tomaron parte en esta investigación, a razón de 4 grupos de 30 escolares cada uno.

Instrumentos y materiales

Se utilizaron los siguientes instrumentos de medida:

«*Self-Perception Profile for Adolescents*» (Harter, 1988), que mide la competencia percibida escolar, aceptación social de los compañeros, competencia atlética, apariencia física externa, competencia en el trabajo, atracción romántica, comportamiento adecuado o conducta, amigos íntimos, la autoestima global, y una puntuación en discrepancia (media de las competencias percibidas menos la importancia concedida a las mismas). La fiabilidad del instrumento oscila entre .74 y .91.

«*Social Support Scale for Children*» (Harter, 1985 c), que mide el apoyo afectivo-social proveniente de los padres, los compañeros de clase, los profesores y los amigos íntimos, siendo el apoyo total la media de los apoyos de las cuatro fuentes. La fiabilidad del instrumento oscila entre .72 y .88, en las diferentes muestras.

«*A Scale of Intrinsic Versus Extrinsic Orientation in the Classroom*» (Harter, 1980), que mide la aproximación a un polo intrínseco vs. extrínseco según los siguientes criterios: a) preferencia por el desafío y el trabajo serio en la clase versus la preferencia por el trabajo fácil guiado únicamente por el profesor. b) Curiosidad e interés del alumno que intenta satisfacer sus propios deseos y necesidades versus el trabajo del mismo que es solamente dirigido a satisfacer al profesor y obtener la aprobación mínima del mismo. c) Independencia del alumno respecto al profesor en cuestiones académicas relevantes que implican participación activa versus dependencia del profesor quien dirige y guía el trabajo del alumno. d) Juicio independiente del alumno capaz de hacer ciertos juicios sobre lo que está haciendo versus conexión con el juicio del profesor de cuya opinión y juicio depende la actuación del alumno. e) Criterio Interno del alumno que implica que éste conoce bien cuando tiene éxito o fracasa en las actividades escolares relevantes versus Criterio Externo que implica dependencia del alumno de fuentes de evaluación y de reforzamiento provenientes del profesor en aspectos de ejecuciones correctas, exámenes y feedback sobre las actividades planteadas en la clase. La fiabilidad de las distintas subescalas (Kuder-Richardson Formula 20) oscila entre 0.68 y .84. La validez factorial arroja índices entre .46 y .53, con insignificantes cargas cruzadas entre los factores.

«*Test de Factor «g» de Cattell*». TEA S.A. (1989), test colectivo que mide la inteligencia general en sus aspectos más generales y cristalizados.

«*Inventario de hábitos de Estudio*» de F.F. Pozar. TEA Ediciones S.A. (1983), que mide aspectos supuestamente relacionados con la eficacia en el estudio: Ambiente, Planificación, Materiales, y Asimilación.

«*Número medio de horas de estudio diarias*» operativizada mediante una cifra que admitía la entrada de hasta dos decimales.

«*Calificación obtenida en las materias académicas*» propias de 4º curso de la ESO, operativizadas mediante un número entero de 0 a 10. Las materias evaluadas fueron las siguientes: Ciencias Sociales, Ética, Educación Física, Lengua y Literatura, Lengua Extranjera, Matemáticas, Biología, Física y Química, y Media de materias optativas.

«*Rendimiento académico*» obtenido en la primera, segunda y tercera evaluación, siendo esta última la variable dependiente para nosotros más relevante, denominada rendimiento escolar final. La operativización de las mismas se hizo de la siguiente forma: Suma de las puntuaciones obtenidas en todas las materias dando un peso de 1 punto a Insuficiente, 5 puntos a Aprobado, 6 puntos a Bien, 7 a Notable y 9 a Sobresaliente.

Variables

El número total de variables que se contabilizan según la descripción anterior es de treinta y nueve que sumadas a la variable Sexo, hacen un total de cuarenta. Las variables dependientes principales fueron dos: la autoestima global (AEG), y el rendimiento final en la tercera evaluación (RENFINAL). Todas las demás variables se seleccionaron en la mayoría de los análisis como variables independientes.

Procedimiento

Durante el curso académico 97-98 se obtuvieron todos los datos relativos a las variables independientes y dependientes. La recogida de datos fue siempre de forma colectiva pero separadamente en cada una de las clases, a pesar de que los sujetos seleccionados de todas ellas se han considerado como un único grupo.

IV. ANÁLISIS DE DATOS Y DISCUSIÓN

Técnicas de análisis de datos

Las técnicas de análisis de datos utilizadas en esta investigación han sido correlacionales y basadas en el BMDP. (BMDP2D, 2R, 6R y 9R). (Dixon y Jennrich, 1990).

Análisis de datos y discusión

Comenzaremos con las correlaciones entre las variables independientes y la variable dependiente autoestima global (AEG). (BMDP2R y 6R). (Tabla 1).

TABLA 1
ANÁLISIS DE REGRESIÓN "STEPWISE" DE LAS PRINCIPALES VARIABLES CON LA VARIABLE DEPENDIENTE AUTOESTIMA GLOBAL EN 4º DE ESO

R Múltiple 0.7528
R Multiple al cuadrado 0.5668
R Ajustada al cuadrado 0.5435
Error std. de est 0.4535

ANÁLISIS DE VARIANZA

	Suma de Cuadrados	Grados de libertad	Media Cuadrática	F Ratio
REGRESIÓN	15.069750	3	5.023249	24.42
RESIDUAL	11.519590	116	0.2057070	

TABLA RESUMEN

Paso Nº	Entrada de variable	Múltiple R	RSQ	Cambio en RSQ	F de Entrada
1	APARIENCIA EXTERNA	0.6721	0.4517	0.4517	47.79
2	APOYO TOTAL	0.7268	0.5282	0.0765	9.24
3	DISCREPANCIA	0.7528	0.5668	0.0385	4.98

Del análisis de esta tabla se deduce que las variables con mayor peso para predecir la variable dependiente Autoestima Global son la *Apariencia física Externa* en primer lugar, seguida del *apoyo total* de padres, compañeros, profesores y amigos íntimos, y finalmente, la *discrepancia media* en las dimensiones que el adolescente se percibe competente en función de la importancia concedida a las mismas.

Es interesante comentar la importancia que en esta edad adquiere la **apariencia y el atractivo físico** de los adolescentes, debido a la presencia de factores hormonales y de crecimiento físico que influyen en las relaciones interpersonales entre los sexos. Es

sorprendente la coincidencia encontrada con los hallazgos de Harter (1988 a) y otros, quienes han encontrado que la apariencia física percibida encabeza repetidamente la lista como la dimensión más altamente correlacionada con la autoestima (r entre .65 y .82) desde la infancia hasta la edad adulta, sin constatarse diferencias de género en dichas correlaciones. Otros autores que han examinado estas relaciones en niños y adolescentes han encontrado resultados similares (Padin, Lerner & Spiro, 1981; Pomerantz, 1979). Además, esta correlación se mantiene igualmente alta en poblaciones de alumnos superdotados y con dificultades de aprendizaje. (Harter, 1993).

Existen numerosos trabajos empíricos relativos al autoconcepto y autoestima adulta que revelan que la **apariencia física percibida** y la autoestima están estrechamente unidas (Davies & Furnham, 1986; Feingold, 1992; Longo y Ashmore, 1995; Silverstein, Striegel-Moore, Timko & Rodin, 1988). No obstante, se ha encontrado que las autoevaluaciones están más altamente relacionadas con el «atractivo percibido» que con el atrativo físico actual, más «objetivo». (Feingold, 1992).

Este aspecto, evidentemente, también puede estar en la base de algunos trastornos psicopatológicos relacionados como la anorexia y labulimia nerviosa, la depresión, ansiedad y suicidio en la adolescencia, etc. (Harter y Marold, 1991), aspectos que tienen especiales implicaciones en el rendimiento académico de los adolescentes y, en consecuencia, en la orientación y la intervención psicopedagógica.

La literatura también arroja datos que apuntan al hecho de que las mujeres suelen estar generalmente más insatisfechas con su apariencia física que los hombres (Adams, 1982; Davies & Furhnam, 1986; Mintz & Betz, 1988). Esta insatisfacción comienza en la infancia media (Maloney, Mc Guire, & Daniels, 1988; Mellin, 1988) declinando progresivamente y siendo la diferencia máxima aproximadamente al final de la educación secundaria obligatoria (Allgood-Merten, Lewinsohn & Hops, 1990; Nolen-Hoeksema, 1987). No obstante, es necesario matizar que la autoestima es particularmente baja en aquellas mujeres que son conscientes de estar basando su propia autoestima sobre el atractivo físico (Harter, 1993). Culturalmente, los mensajes acerca del atractivo y apariencia física comienzan a edades tempranas (Maccoby & Martin, 1983).

Este hecho debe hacernos reflexionar sobre el énfasis excesivo que la sociedad actual y los medios de comunicación ejercen sobre el atractivo físico en cada periodo evolutivo. (Andersen, 1992; Elkind, 1979, Kilbourne, 1994). El cine, la televisión, las revistas, el vídeo, etc., enfatizan la importancia del culto al cuerpo enalteciendo el papel popular de modelos a quienes hombres y mujeres deberían emular. No obstante, estos estándares son especialmente nocivos para las mujeres (Jackson, 1992; Wiseman, Gray, Mosimann & Ahrens, 1992), haciendo difícil vivir de acuerdo a estos ideales.

Este aspecto, evidentemente, también puede estar en la base de algunos trastornos psicopatológicos relacionados como la anorexia, bulimia, depresión, ansiedad y suicidio en la adolescencia, etc., (Harter y Marold, 1991a), aspectos que tienen especiales implicaciones para la intervención.

Las correlaciones parciales halladas entre la autoestima global con las demás variables más relevantes se presentan a continuación: Apariencia física externa (r=.67); Discrepancia (r=.58); apoyo de compañeros (r=.45); apoyo de amigos íntimos (r=.40);

atracción romántica y competencia atlética (r=.37); apoyo de padres y competencia escolar (r=.35); rendimiento en la 1ª evaluación (r=-.022); rendimiento en la 2ª evaluación (r=.009) y rendimiento a final de curso (r=.147).

De los datos anteriores puede concluirse que la apariencia física externa, la discrepancia y el apoyo total por compañeros, padres y amigos, así como el apoyo de compañeros/as con quienes se tiene una relación más íntima y afectivo-sexual parecen contribuir en mayor medida sobre la autoestima de los adolescentes. Estos resultados están en consonancia con los estudios realizados por Harter en la población americana. Se han incluido las bajas correlaciones encontradas con el rendimiento escolar en las tres evaluaciones del curso para llamar la atención sobre este hecho, que parece coincidir con la observación y los datos aportados por las entrevistas mantenidas con padres y alumnos de enseñanza secundaria obligatoria. Con estos datos se pretende constatar, todavía en un plano especulativo, que el alto o bajo rendimiento escolar en la población actual de estudiantes de secundaria, podría contribuir poco al desarrollo de un adecuado sentimiento de autoestima y que los adolescentes, parecen abastecerse, por decirlo así, de fuentes independientes del aprovechamiento escolar, o al menos, no totalmente coincidentes. Se podría llegar a afirmar que muchos alumnos con bajo rendimiento escolar pueden llegar a desarrollar altos niveles de autoestima si desarrollan adecuadamente las dimensiones comentadas más arriba o, por el contrario, mantener una baja autoestima a pesar de conseguir un alto rendimiento académico. La alta autoestima global y el alto rendimiento académico podrían tener dimensiones causales distintas y desarrollarse por caminos no totalmente convergentes. No obstante, este aspecto debería investigarse con más profundidad en subgrupos en los que la dimensión «competencia escolar» es especialmente alta o baja, teniendo en cuenta que se observan puntuaciones en importancia igualmente altas o bajas, lo que puede servir de entrada a otras facetas o dimensiones «más significativas» y a las que conceden mayor importancia.

Pero vayamos ahora a analizar la influencia relativa que sobre el rendimiento en la evaluación final tienen las variables independientes seleccionadas. (Tabla 2).

Esta tabla revela que variables, hasta cierto punto inesperadas, entran como predictoras, como son la inteligencia y las horas medias diarias dedicadas al estudio, así como algunas materias académicas troncales, que en líneas generales, constituyen la parte básica del currículum de este curso.

La variable **inteligencia** ha sido puesta de manifiesto en publicaciones de psicología o pedagogía diferencial como variable predictora a tener en cuenta en relación al rendimiento escolar, como por ejemplo los trabajos de Eysenck y Kamin (1983), o Jiménez Fernández (1997), entre otros.

En relación a las **horas de estudio** (como segunda variable predictora), cabría decir que el tiempo dedicado al mismo, (no superior a una media de 2 horas diarias) puede ser una variable actualmente relevante en el rendimiento escolar, y esto nos llevaría a reactualizar el modelo de Carroll (1963), el cual está siendo redescubierto recientemente por el significado e impacto de la variable tiempo en el proceso educativo y en sus resultados. Este hecho podría estar indicando una insuficiente inversión de tiempo de estudio por parte de algunos alumnos.

TABLA 2
El «MEJOR» SUBCONJUNTO DE REGRESIÓN DE VARIABLES INDEPENDIENTES SOBRE EL RENDIMIENTO ACADÉMICO FINAL EN 4º DE ESO. (BMDP9R)

ESTADÍSTICOS PARA EL "MEJOR" SUBCONJUNTO DE REGRESIÓN DE VARIABLES INDEPENDIENTES SOBRE RENDIMIENTO FINAL

Coef. de Mallows 1.64
Correlación múltiple al cuadrado 0.96325
Correlación múltiple 0.98145
Corr. múltip. al cuadrado ajustada 0.95575
Media cuadrática residual 6.767583
Error estándar de estimación 2.601458
Estadístico F 28.43
Grados de libertad del numerador 10
Grados de libertad del denominador 49
Nivel de significación 0.0000

VAR.	COEFI. REGRES	ERROR STD.	COEFI. STD.	ESTAD. T	NSIG. 2-TAIL	TOLER.	CONTRIB. A RSQ.
Intercept	-7.66082	2.15081	-0.619	-3.56	0.001		
Inteligencia	0.03881	0.01521	0.078	2.55	0.014	0.8079	0.00488
Horas est.	0.36608	0.19314	0.061	1.90	0.064	0.7232	0.00269
Ambiente	0.58557	0.22845	0.088	2.56	0.013	0.6388	0.00493
CSociales	0.85921	0.32381	0.117	2.65	0.011	0.3889	0.00528
Lengua y L.	1.56228	0.35907	0.187	4.35	0.000	0.4077	0.01420
Matemat.	2.06921	0.35468	0.244	5.83	0.000	0.4271	0.02553
FQuímica	2.18923	0.34979	0.259	6.26	0.000	0.4383	0.02938
Moptativ.	1.38392	0.33843	0.147	4.09	0.000	0.5808	0.01254

La tercera variable (**ambiente de estudio**) puede introducirnos en un campo todavía muy extenso y poco delimitado, en el que pueden entrar a formar parte factores como clima familiar (Dueñas Buey, 1988), clase social y educación (Minton y Schneider, 1980; Lautrey, 1985) o los estudios relativos a medio rural, urbano suburbano y educación, que tienden a constatar diferencias que favorecen a las clases socialmente más elevadas y al medio urbano, entre otras variables igualmente relevantes (Consejo de Universidades, 1987; Martin, 1985; MEC-INCE, 1998).

Finalmente, en lo que respecta a las variables de **materias académicas** propiamente dichas que predicen, después de las variables anteriores, el rendimiento académico, encontramos las materias de Ciencias Sociales, Lengua y Literatura, Matemáticas y Física y Química. Teniendo en cuenta el Informe emitido por el MEC de las estadísticas en dicho Territorio de gestión (tanto centros públicos como privados) y relativas a

la promoción de alumnos de 4º de ESO en el curso académico 1996-97, detectamos que el número de alumnos que superan positivamente las materias comentadas anteriormente son, precisamente las cuatro más bajas, junto con el idioma extranjero que menor porcentaje de alumnos supera a final de curso (Física y Química: 77, 7 % de evaluación positiva; Ciencias Sociales: 77,6 evaluados positivamente; Lengua y Literatura: 72,9% de evaluación positiva; Matemáticas: 66,1% de evaluación positiva; e idioma extranjero (inglés): 68,5% de aprobados.

En resumen, aproximadamente un alumno de cada cuatro en dichas materias (y uno de cada tres en Matemáticas) no alcanza los objetivos mínimos para superar la materia correspondiente a final de curso, lo cual puede hacer suponer que ya no son factores únicamente intrapersonales del alumno sino igualmente pedagógicos (estilos de enseñanza, metodología, variables de profesor como, por ejemplo, la formación, interacción alumno-profesor, organización del aula, estilos de enseñanza, inadecuación de los criterios de evaluación, etc.), o institucionales (el propio sistema educativo), las que pueden estar implicadas en dicho aprovechamiento escolar.

Eliminando el efecto de las las materias académicas, los resultados aparecen de la siguiente forma. (Tabla 3).

TABLA 3

EL "MEJOR" SUBCONJUNTO DE REGRESIÓN DE VARIABLES INDEPENDIENTES SOBRE EL RENDIMIENTO ACADÉMICO FINAL EN 4º DE E.S.0. DESPUÉS DE ELIMINAR LAS MATERIAS ACADÉMICAS. (BMDP9R)

ESTADÍSTICOS PARA EL "MEJOR" SUBCONJUNTO DE REGRESIÓN DE VARIABLES INDEPENDIENTES SOBRE RENFINAL

Correlación múltiple al cuadrado 0.84271
Correlación múltiple 0.91799
Corr. múltip. al cuadrado ajustada ..... 0.828[illegible]5
Error estándar de estimación 5.126549
Estadístico F 57.86
Nivel de significación 0.0000

VAR.	COEFI. REGRES	ERROR STD.	COEFI. STD.	ESTAD. T	NSIG. 2-TAIL	TOLER.	CONTR. A RSQ.
Intercept	9.49839	3.03195	0.768	3.13	0.003		
Intern/ext	2.71942	1.01744	0.173	2.67	0.010	0.695655	0.02081
Ambiente	1.07138	0.46059	0.161	2.33	0.024	0.610318	0.01576
Material	0.64960	0.37729	0.111	1.72	0.091	0.694998	0.00863
Renduno	0.28954	0.12234	0.310	2.37	0.022	0.169263	0.01631
Rendos	0.32231	0.11159	0.381	2.89	0.006	0.167344	0.02430

La variable predictora con más peso es la dimensión **interno/externo**, es decir, en qué medida el alumno atribuye el éxito o el fracaso a factores internos y a la propia autorregulación de su aprendizaje, o por el contrario a factores externos, de azar o suerte. Esta variable puede tener implicaciones importantes para la intervención educativa y la tutoría por sus relaciones con la motivación académica, pero aunque este aspecto ha cobrado especial importancia en los últimos años (Harter, 1981, 1984, 1992; Harter & Jackson, 1992; Harter, Whitesell & Junkin, 1997; Harter, Whitesell & Kowalski, 1992; González Torres, 1997; Alonso Tapia, 1998), y otros enfoques recientes comentados más arriba, creemos que su aplicación práctica en el aula es practicamente inexistente por parte de los docentes en la actualidad.

Por otra parte, además del **ambiente y el material**, la **evaluación obtenida en la primera y segunda evaluación** se convierten en predictores importantes del rendimiento escolar final, apuntando al hecho de que, en la medida en que se comienza y se mantiene el alumno a lo largo del curso, es muy probable que finalice el mismo con o sin éxito escolar. No obstante, este aspecto podría tener explicaciones alternativas de especial relevancia en relación con el tema actual del problema de la evaluación en la educación secundaria. El posible determinismo y continuidad en la evaluación por parte del profesorado que empieza a observarse en relación con alumnos y alumnas de la ESO a lo largo del curso ¿significa realmente una consistencia por parte de éstos en la adecuada o inadecuada adquisición y mantenimiento de los contenidos, en sus hábitos de estudio, en sus aptitudes y rendimiento académico en un curso determinado?, o por el contrario ¿implica una perpetuación sesgada en los métodos de evaluación por parte del profesorado y en una percepción y catalogación previa del alumno, desde el principio, que impide al profesor/a detectar fluctuaciones en el aprendizaje, necesidades educativas transitorias o adelantos significativos en momentos específicos de la historia de aprendizaje de aquel? o finalmente, ¿es que los profesores desarrollan de forma especial un sentido intuitivo para pronosticar de forma infalible, y casi a priori, el destino académico de sus alumnos? Estas y otras razones suponen un reto para la investigación futura de los profesionales de la educación interesados en el tema.

V. CONCLUSIONES Y PERSPECTIVAS FUTURAS

Los resultados obtenidos en esta investigación empírica confirman y están en consonancia con el modelo teórico cognitivo-evolutivo de S. Harter, al menos en el rango de edad estudiado (16-17 años) y en la muestra española. Este modelo parece uno de los que más se acerca a los datos empíricos en lo que se refiere a aspectos de desarrollo evolutivo y especificación de los determinantes de la autoestima, así como en aspectos parciales relacionados con el logro escolar.

De acuerdo a lo que se ha constatado y se ha expuesto en este trabajo, el *atractivo físico percibido* se encuentra estrechamente unido a la autoestima y a la depresión. Las intervenciones en este campo son, en la actualidad, verdaderamente desafiantes. Pope et al. (1988) ofrecen una serie de sugerencias. En la línea de la literatura *atribucional* estos autores creen que es prioritario el cambio de las atribuciones individuales relati-

vas a la falta de atractivo físico. Aunque sugieren que algunas orientaciones centradas en la forma de vestir y en la higiene personal pueden ser apropiadas, no recomiendan soluciones de «cosmética y estética» que tanto han sido asociadas a los ideales de atractivo físico puestas de manifiesto constantemente en los medios de comunicación social. Más bien, sugieren que los agentes educativos y familiares transmitan la idea de que tales ideales son virtualmente inalcanzables y que las personas deberían alterar y cambiar los estándares sociales a los que a ellos les gustaría parecerse. Desde la perspectiva de James, lo más sugerente es reducir la discrepancia entre la imagen ideal del propio cuerpo y la percepción real del mismo. Pope et al., también recomiendan algún tipo de *reestructuración cognitiva* que modifique el centro de atención sobre el propio atrativo físico hacia otras dimensiones igualmente relevantes en los individuos como pueden ser la académica, atlética, interpersonal o moral en las cuales estos individuos son, o pueden llegar a conseguir, mayor éxito, es decir, los esfuerzos de intervención deberían dirigirse al cambio de percepción en los individuos de que las cualidades o características que configuran su «yo interno» como persona valiosa determinarán inevitablemente la aceptación de su «yo externo».

Dentro del *contexto escolar*, se han diseñado programas educativos para mantener o intensificar el autoconcepto académico y la autoestima global. Los objetivos de estos diversos programas reflejan dos orientaciones que parten de presupuestos teóricos que compiten entre sí. Como han subrayado Caslyn & Kenny (1977), los **teóricos del fortalecimiento del yo («self-enhancement»)** creen que los esfuerzos de intervención deben dirigirse a realzar el autoconcepto y la autoestima directamente, por ejemplo, dando a los estudiantes afectividad y apoyo basándose en ejercicios que estimulen y fortalezcan, en líneas generales, los sentimientos de «sí mismos». Por el contrario, los **teóricos del «desarrollo de destrezas y habilidades»** argumentan que las actitudes acerca del «yo» son consecuencias del logro académico y por tanto, los esfuerzos pedagógicos deberían dirigirse hacia el logro de destrezas académicas específicas. En los últimos años, la balanza se ha inclinado hacia la orientación del aprendizaje de habilidades cuyas intervenciones se han dirigido hacia dimensiones específicas.

Además de ello, recientemente (Harter, 1998), ha sugerido la utilidad de distinguir entre el objetivo de un programa (por ejemplo, fortalecer la autoestima) con lo que podría denominarse «el blanco» de las intervenciones, argumentando que mientras el fortalecimiento de la autoestima puede ser una meta, las estrategias de intervención deberían dirigirse hacia sus **determinantes**. Dentro de este modelo, las intervenciones para fortalecer la autoestima deberían intentar reducir la discrepancia entre las dimensiones de competencia/importancia (que en este trabajo aparecen como una de las principales variables predictoras), y de forma complementaria, encontrar formas de incrementar la aprobación de los «otros significativos», proporcionando apoyo de agentes familiares o sociales de los cuales los adolescentes pueden obtener una visión más positiva. En este sentido, los programas de la década de los años 60 y 70 que intentaban aumentar la autoestima del jóven directamente tuvieron poco impacto sobre la misma. Otros investigadores han ofrecido argumentos similares (Greenberg et al. 1995), centrándose prioritariamente sobre el aprendizaje de destrezas. Hattie (1992) en un meta-análisis realizado sobre una variedad de programas de intervención

concluye que los programas cognitivos son consistentemente y de forma significativa más efectivos que los programas basados unicamente en lo afectivo (ver también Strein, 1988). Hattie sugiere que las intervenciones orientadas hacia lo *cognitivo*, con objetivos más específicos y definidos son también más susceptibles de valoración. Marsh (Marsh & Hattie, 1996) afirma que esta perspectiva es mucho más aconsejable. En este sentido, se ha constatado que las intervenciones académicas tienen efectos significativos sobre los componentes académicos del autoconcepto pero muy poco efecto sobre los componentes no académicos y viceversa. Bracken (1996) también ha señalado que la carencia de éxito de ciertos programas de intervención se debe al hecho de que el autoconcepto global es insuficientemente sensible a tratamientos específicos, recomendando que las intervenciones se dirijan directamente hacia varios componentes del autoconcepto *simultáneamente*. En una línea parecida se posicionan autores como Damon (1995), Seligman (1993) y Broc (1991).

Por otra parte, y en otro nivel de análisis, partiendo de que un adolescente tiene una baja autoestima ello no nos permite todavía predecir a qué tipo específico de conducta o trastorno dará lugar (fracaso escolar, depresión, suicidio, trastornos de la conducta alimentaria, conductas antisociales y/o delincuencia, etc.). Por lo tanto, el desafío actual es desarrollar modelos que identifiquen antecedentes específicos de diferentes trastornos posteriores que mantengan el papel de las auto-representaciones como mediadores fenomenológicos válidos. (Harter, 1998).

Adentrándonos en intervenciones más específicas, Hattie (1992) sugiere que proporcionando al alumno *expectativas realistas* que sean ligeramente más elevadas que las que mantiene actualmente se podría fortalecer el sentido de auto-eficacia en dominios relevantes del autoconcepto (cf., Bandura, 1990). Otros autores se han centrado básicamente sobre aquellas variables *atribucionales* del sentido individual del «yo» y que están implicadas en trastornos depresivos. De esta forma, estrategias en las que se trabaja una *reestructuración cognitiva-verbal* de las propias atribuciones modificándolas en internas, estables y globales en relación al propio logro personal se han sugerido como nuevas estrategias de intervención (Pope, McHale, & Craighead, 1988; Seligman, 1993).

Se han recomendado otros tipos de intervención a lo largo del continuum de programas que van desde lo educacional hasta las terapias más formales. A modo de síntesis, podría afirmarse que en el extremo más «educativo» Beane (1994) describe la *aproximación «ecológica»* como el modelo por el que se intentan alterar características del contexto escolar más amplio de tal forma que proporcione una atmósfera más óptima y funcional para el desarrollo del autoconcepto y la autoestima (por ejemplo, cambiando el énfasis de un control externo hacia otras formas de autocontrol más personal). Otras aproximaciones terapéuticas más específicas se han basado en la terapia cognitiva (Beck, 1976; Seligman, 1993), la terapia racional emotiva (Ellis, 1958; Zastrow, 1994) y en intervenciones psicoanalíticas (Wexler, 1991) basadas en la formulación de Kohut (1977).

Es necesario subrayar que un problema principal relacionado con los esfuerzos puestos en el diseño de programas de intervención ha sido el fracaso en emplear una estrategia de evaluación adecuada en cada programa. Algunas estrategias de eva-

luación propuestas para mejorar la valoración adecuada de dichos programas podrían ser las siguientes: a) seleccionar instrumentos que de forma específica midan los constructos que constituyen el objetivo de la intervención y que tengan fundamentadas propiedades psicométricas; b) intentar especificar un *modelo de predicciones*, que indiquen qué tipo de trastornos o «productos conductuales» serían o no serían afectados, y que incluyeran medidas fiables y válidas de cada uno de ellos (ver Marsh & Hattie, 1996); c) incluir medidas de los procesos actuales que supuestamente son responsables del cambio en el autoconcepto; d) más que esperar ganancias medias a nivel de grupo, identificar subgrupos que puedan y/o no puedan aprovecharse de la intervención, intentando identificar igualmente los factores que pueden dar lugar a estos diferentes tipos de resultados en la intervención. Es posible que existan adecuados programas de intervención cuya eficacia no haya podido ser demostrada debido al problema de incluir una evaluación adecuada y estrategias adecuadas de análisis de datos.

A pesar de los relativos avances comentados en este trabajo todavía quedan desafíos importantes para la futura investigación en este campo. Algunos de los análisis evolutivos son todavía excesivamente teóricos y descriptivos, y carecen de especificidad así como de un fundamento empírico sólido, particularmente en lo que respecta a los factores subyacentes y reglas de transición que expliquen los mecanismos del cambio. Aunque los investigadores de los procesos del «yo adulto» han conseguido un nivel más amplio de especificidad, sabemos muy poco acerca de cómo dichos procesos emergen evolutivamente. A lo largo de todos los niveles evolutivos, se debería de prestar mucha más atención a variables contextuales que ejercen un impacto decisivo en las auto-representaciones. Finalmente, se necesita tener mucho cuidado en la investigación acerca del «yo», es decir, es necesario explicar cómo las auto-representaciones tienen consecuencias críticas en la vida diaria de los niños, adolescentes y adultos, para que los esfuerzos de los investigadores y educadores no vayan por caminos equivocados. Todos ellos son desafios estimulantes que se nos presentan para el futuro.

BIBLIOGRAFÍA

ADAMS, G.R. (1982). Physical attractiveness. In A.G. Miller (Ed.), *In the eye of the beholder: Contemporary issues in stereotyping* (pp. 54-79). New York: Praeger.

ALONSO TAPIA, J. (1992 a). *Motivar en la adolescencia: Teoría, evaluación e intervención.* Madrid: Ediciones Universidad Autónoma.

ALONSO TAPIA, J. (1992 b). *¿Qué es lo mejor para motivar a mis alumnos? Análisis de lo que los profesores saben, creen y hacen al respecto.* Colección cuadernos del ICE, nº 5, Madrid: Servicio de Publicaciones de la Universidad Autónoma.

ALONSO TAPIA, J. (1995). *Orientación Educativa.* Madrid: Sintesis.

ALONSO TAPIA, A. (1998). *Motivación y aprendizaje en el aula. Cómo enseñar a pensar.* Aula XXI. Santillana.

ALLGOOD-MERTEN, B., LEWINSOHN, P.M., & HOPS, R. (1990). Sex differences and adolescent depression. *Journal of Abnormal Psychology, 99*, 55-63.

ANDERSEN, A.E. (1992). Diet vs. shape content of popular male and female magazines: A dose-response relationship to the incidence of eating disorders? *International Journal of Eating Disorders, 11* (3), 283-287.

BANDURA, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2), 191-215.

BANDURA, A. (1990). Conclusion: Reflections on nonability determinants of competence. In R. J. Sternberg & J. Kolligian, Jr. (Eds.), *Competence considered* (pp. 316-352). New Haven, CT: Yale University Press.

BEANE, J.A. (1994). Cluttered terrain: The schools´ interest in the self. In T. M. Brinthaupt & R. P. Lipka (Eds.) *Changing the self* (pp. 69-88). Albany: State University of New York Press.

BECK, A.T. (1976). *Cognitive therapy and the emotional disorders*. New York: New American Library.

BRACKEN, B. (1996). Clinical applications of a context-dependent multi-dimensional model jof self-concept. In B. Bracken (Ed.), *Handbook of self-concept* (pp. 463-505). New York: Wiley.

BROC, M.A. (1994). Rendimiento académico y autoconcepto en la Educación Infantil y Primaria. En *Revista de Educación, nº 303*, 281-297. Madrid. C.I.D.E.

BROC, M.A. (1995). Evaluación e Intervención en el autoconcepto y autoestima en contextos educativos. *Revista de Psicología Educativa. (Revista de los psicólogos de la Educación del Colegio Oficial de Psicólogos). Vol. 1. Nº 1*, 53-62.

BROC, M.A. (1996). Orientaciones de intervención dirigidas al profesorado para favorecer el desarrollo del autoconcepto en la educación infantil, primaria y secundaria. *IberPsicologia. Anales de la «Revista de Psicología General y Aplicada» Vol. 1.* En INTERNET, en la localización *Iberpsicología, 1996:1.1.9* correspondiente a http://fs-morente.filol.ucm.es/ publicaciones/Iberpsicología /Iberpsicología.htm/ Iberpsi1/Broc/Broc.htm

BROC, M.A. (1998). *Influencia relativa de cuatro antecedentes de la autoestima e intervención experimental en niños de 5 a 8 años de edad*. Tesis doctoral publicada en microficha en el Servicio de Investigación de la UNED.

BROC, M.A. (2000). Procesos que fundamentan la formación del autoconcepto en la adolescencia. En *ANNALES*. Anuario de la UNED de Barbastro (Huesca). (En prensa).

BROC, M.A., y GARCÍA, B. (1993). Cinco programas de intervención sobre dimensiones del autoconcepto. (I Congreso Internacional de Psicología y Educación. Intervención Psicoeducativa). Dto. de Psicología Evolutiva y de la Educación. En *Líneas actuales en la intervención psicopedagógica II. Variables personales y psicosociales. Vol. II.*, 811-822. Editorial Complutense. Madrid.

BROC, M.A. & HUGUET, A. (1996). Programas cognitivo-experimentales para la mejora del rendimiento académico. En *Educación Cognitiva*. Vol. II, 71-79. Santiago Molina y Manuel Fandos (Coordinadores). Mira Editores.

BROC, M.A. & HUGUET, A. (1997). Programas para favorecer el desarrollo del autoconcepto en contextos de enseñanza-aprendizaje. Un estudio longitudinal. En Beltrán y otros. *Nuevas perspectivas en la intervención psicopedagógica: I. Aspectos*

140

cognitivos, motivacionales y contextuales. 223-226. Dto. de Psicología Evolutiva y de la Educación. Universidad Complutense de Madrid. Editorial Complutense.

CARROLL, J. B. (1963). «A model of school learning», en *Teachers College Record*, (64), 723-733.

CASLYN, R.J., & KENNY, D.A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology, 69*, 136-145.

CATTELL, (1989). *Test de factor «g»*. TEA Ediciones S.A. Madrid.

CONNELL, J.P. (1990). Context, Self and Action: A motivational Analysis of Self-System Processes across the Life Span. En Cicchetti, D.; Beegly, M. (Eds.). *The Self In Transition: Infancy to Childhood*, pp. 61-98. Chicago: The University of Chicago Press.

CONSEJO DE UNIVERSIDADES. (1987). *Demanda de plazas universitarias. MEC. Madrid.*

COOLEY, C.H. (1902). *Human nature and the social order*. New York: Charles Scribner´s Sons.

COOPERSMITH, S. (1967). *The antecedents of self-esteem*. San Francisco, CA: W.H. Freeman & Co.

COVINGTON, M.V. (1984). The Motive for Self-Worth. En Ames, R.; Ames, C. (Eds.). *Research on Motivation in Education*. V.I.: *Student Motivation*, 78-114. Academic Press. Boston.

CSIKSZENTMIHALYA, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.

DAMON, W. (1995). *Greater expectations: Overcoming the culture of indulgence in America´s homes and schools*. New York: Free Press.

DAMON, W., & HART, D. (1988). *Self-understanding in childhood and adolescence*. Cambridge University Press.

DÁVIES, E., & FURNHAM, A. (1986). Body satisfaction in adolescent girls. *British Journal of Medical Psychology, 59*, 279-287.

DECI, E.L. & RYAN, R.M. (1992). The Iniciationand Regulation of Intrisically Motivated Learning and Achievement. En Boggiano, A.K.; Pittman, T.S. (Eds.). *Achievement and Motivation: A Social-Developmental Perspective*. New York: Cambridge University Press.

de CHARMS, R. (1994). Motivation Enhancement in Educational Settings. En Ames, R.; Ames, C. *Research on Motivation in Education*. V.I: *Student Motivation*, pp. 275-312. Boston: Academic Press.

DIXON, W.J., & JENNRICH, R. (1990). Stepwise Regression. Págs. 359-394. En *BMDP. Statistical Software. Manual. Vol. 1*. University Press of California. Berkeley. Los Angeles. Oxford.

DWECK, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist*, 41 (10), 1040-1048.

DUEÑAS BUEY, M.L. (1988). *Análisis y valoración de un ensayo de integración escolar en Madrid capital*. Tesis Doctoral. Dto. MIDE. Facultad de Filosofía y Ciencias de la Educación. Madrid. U.N.E.D.

ECCLES, J.S. & MIDGLEY, C. (1989). Stage/environment fit: Developmentally appropiate classrooms for early adolescents. En R. Ames & C. Ames (Eds.). *Research on motivation in education*. (Vol. 3, pp. 139-181). San Diego, CA: Academic Press.

ELKIND, D. (1979). Growing up faster. *Psychology Today, 12,* 38-45.

ELLIS, A. (1958). Rational Psychotherapy. *Journal of General Psychology, 58,* 35-49.

EYSENCK, H.J. & KAMIN, L. (1983). *La confrontración sobre la inteligencia. ¿Herencia o ambiente?* Madrid: Pirámide.

FEINGOLD, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304-341.

GARCÍA, B. (1983). *Análisis y delimitación del constructo autoestima.* Tesis Doctoral presentada en la Universidad Complutense de Madrid.

GONZÁLEZ, M. C., TOURON, J. (1992) *Autoconcepto y Rendimiento Escolar. Sus Implicaciones en la Motivación y en la Autorregulación del Aprendizaje.* Pamplona: Eunsa.

GONZÁLEZ, M.C. (1997). *La motivación académica. Sus determinantes y pautas de intervención.* EUNSA. Ediciones de la Universidad de Navarra. Pamplona.

GREENBERG, J., PYSZCZYNSKI, T., & SOLOMON, S. (1995). Toward a dual-motive depth psychology of self and social behavior. In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 73-101). New York: Plenum Press.

HARTER, S. (1980). *A Scale of Intrinsic Versus Extrinsic Orientation in the Classroom.* Universidad de Denver. Colorado. EEUU.

HARTER, S. (1981). A model of intrinsic mastery motivation in children: Individual differences and developmental change. *Minnesota Symposium on Child Psychology,* Vol. 14. Lawrence Erlbaum, 1981.

HARTER, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developments Psychology,* 1981, 17, 300-312.

HARTER, S. (1983). Developmental perspectives on the self-system. En M. Hetherington (Ed.). *Handbook of child psychology: Social and personality development. Vol. 4.* New York: Wiley.

HARTER, S. (1985). Competence as a dimension of self-evaluation: Toward a comprehensive model of self-worth. En R. Leahy (Ed.), *The development of the self.* New York. Academic Press.

HARTER, S. (1985 a). Processes underlying the Construction, Maintenance and Enhancement of the Self-Concept in Children. In Jerry Suls and Anthony G. Greenwald (Eds.). *Psychological Perspectives on the Self. Vol. 3.* Lawrence Erlbaum Associates. Hillsdale, New Hersey, London.

HARTER, S. (1985 c). *The Social Support Scale for Children.* Manual. University of Denver. Colorado.

HARTER, S. (1987). Causes, correlates and the functional role functional of global self-worth: A life-span perspective. En J. Kolligian and R. Stenberg (Eds.). *Perceptions of Competence and Incompetence Across the Life-Span.* New Haven, Ct.: Yale University Press.

HARTER, S. (1987). The determinants and Mediational Role of Global Self-Worth in Children. En *Contemporary Topics in Developmental Psychology.* Nancy Eisenberg (Ed.). John Wiley & Sons.

HARTER, S. (1988). *Manual for the Self-Perception Profile for Adolescents.* Universidad de Denver. Colorado.

HARTER, S. (1988 a). Issues in the assessment of the self-concept of children and adolescents. En la Greca, A. (Ed.). *Childhood Assessment: Through the eyes of a child.* Allyn and Bacon.

HARTER, S. (1992). The relationship between perceived affect, and motivational orientation within the classroom: Process and patterns of change. In A. K. Boggiano & T. Pittman (Eds.), *Achievement and motivation: A social-developmental perspective.* Cambridge University Press.

HARTER, S. (1993). Causes and consequences of low self-esteem in children and adolescents. In R.F. Baumeister (Ed.), *Self-Esteem: The puzzle of low self-regard* (pp. 87-116). New York: Plenum Press.

HARTER, S. (1998). The Development of Self-Representations. En W. Damon y N. Eisenberg. *Handbook of Child Psychology.* Vol. 3: *Social, Emotional, and Personality Development.* New York: John Wiley & Sons, Inc.

HARTER, S. & BROC, M. A. (2000). *Perfil de Autoconcepto, Autoestima y Apoyo social para Adolescentes.* Adaptación española. Centro de Sociología y Psicología Aplicadas (COSPA). Madrid. (En prensa).

HARTER, S., & CONNELL, J. P. (1984). A model of children´s achievement and related self-perceptions of competence, control, and motivational orientation. In J. Nicholls (Ed.), *Advances in motivation and achievement, 3,* 219-250. Greenwich, CT: JAI Press.

HARTER, S., & JACKSON, B. (1992). Trait versus nontrait conceptualizations of intrinsic/extrinsic motivational orientation. *Motivation and Emotion, 16,* 209-230.

HARTER, S., & MAROLD, D.B. (1991 a). A model of the determinants and mediational role of self-worth: Implications for adolescent depression and suicidal ideation. In G. Goethals & J. Strauss (Eds.), *The self: An interdisciplinary approach.* Springer-Verlag.

HARTER, S., & PIKE, R. (1984). The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. *Child Development, 55,* 1969-1982.

HARTER, S., WHITESELL, N., & JUNKIN, L. (1997). Similarities and differences in domain specific and global self-evaluations of learning-disabled, behaviorally-disordered and normally-achieving adolescents. (Under review).

HARTER, S., WHITESELL, N., & KOWALSKI, P. (1992). Individual differences in the effects of educational transitions on young adolescents´ perceptions of competence and motivational orientation. *American Education Research Journal, Vol. 29,* pp. 777-807.

HATTIE, J. (1992). *Self-concept.* Hillsdale, NJ: Erlbaum.

HUGHES, H. M. (1984). Measures of self-esteem for prescholl and kindergarten age children: Teacher report mesasures. *Child Study Journal, 14,* 36-41.

JACKSON, L.A. (1992). *Physical appearance and gender: Sociobiological and sociocultural perspectives.* New York: State University of New York Press.

JAMES, W. (1892). *Psychology: The Briefer Course.* New York: Henry Holt & Co.

JIMÉNEZ FERNÁNDEZ, C. (1997). *Pedagogía Diferencial.* Madrid. U.N.E.D.

KILBOURNE, J. (1994). Still killing us softly: Advertising and the obsession with thinness. In P. Fallon, M. Katzman, & S. Wooley (Eds.), *Feminist perspectives on eating disorders* (pp. 395-418). New York: Guilford Press.

KOHUT, H. (1977). *The restoration of the self.* New York: International Universities Press.

LAUTREY, J. (1985). *Clase social, medio familiar e inteligencia.* Infancia y Aprendizaje/ Visor. Madrid.

LEWIS, M., & BROOKS-GUNN J. (1979 b). *Social cognition and the acquisition of self.* New York. Plenum Press.

LONGO, L.C., & ASHMORE, R.D. (1995). The looks-personality relationship: Self-orientations as shared precursors of subject physical attractiveness and self-ascribed traits. *Journal of Applied Social Psychology, 25,* 371-398.

MACCOBY, E., & MARTIN, J. (1983). Socialization in the context of the family: Parent-Child Interaction. En E. H. Hetherington (Ed.), *Handbook of child psychology. Vol. 4: Socialization, personality and social development.* New York: Wiley.

MACHARGO, J. (1991). *El profesor y el autoconcepto de sus alumnos. Teoría y práctica.* Editorial Escuela Española, S.A.

MALONEY, M.J., McGUIRE, J.B., & DANIELS, S.R. (1988). Reliability testing of a children´s version of the Eating Attitude Test. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 541-543.

MARSH, H. W., & HATTIE, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-copncept* (pp. 38-90). New York: Wiley.

MARTIN, T. (1985). *Variables de influjo inmediato en el rendimiento escolar.* Tesis Doctoral. Dto. de Didáctica/U.N.E.D. Madrid.

McCOMBS, B.L. (1992). What Are the Parameters of a New Paradigm of Motivation? *Paper presented in the Symposium: What can a New Paradigm of Motivation Contribute to Practice? Evidence from a Variety of Applications.* Annual meeting of the American Educational Research Association. San Francisco. Abril.

MEC-INCE. (1998). Elementos para un diagnóstico del Sistema Educativo Español. Informe Global. En *Diagnóstico del Sistema Educativo. La Escuela Secundaria Obligatoria.*

MELLIN, L.M. (1988). Responding to disordered eating in children and adolescents. *Nutrition News, 51,* 5-7.

MESSER, B., & HARTER, S. (1989). *The Self-Perception Profile for Adults.* University of Denver. Colorado.

MINTON, H.L. & SCHNEIDER, F.W. (1980). *Differencial Psychology,* Brooks/Cole Publish. Co., Monterrey, California.

MINTZ, L.B., & BETZ, N.E. (1988). Prevalence and correlates of eating dissordered behaviors among undergraduate women. *Journal of Counseling Psychology, 35* (4), 463-471.

NICHOLLS, J.G. (1984 a). Achievement Motivation: Conceptions of Ability, Subjetive Experience, Task Choice, and Performance. *Psychological Review,* 91 (3), 328-346.

NEEMAN, J., & HARTER, S. (1986). *The Self-Perception Profile for College Students. (Manual).* University of Denver. Colorado.

NOLEN-HOEKSEMA, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin, 101,* 259-282.

NÚÑEZ PÉREZ, J.C.; GONZÁLEZ-PIENDA, J.A. (1994). *Determinantes del rendimiento académico. Variables cognitivo-motivacionales, atribuciones, uso de estrategias y autoconcepto.* Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

PADIN, M.A., LERNER, R.M., & SPIRO, A. (1981). Stability of body attitudes and self-esteem in late adolescence. *Adolescence, 62,* 371-384.

PARIS, S.G. & TURNER, J.C. (1994). Situated Motivation. En Pintrich, P.R.; Brown, D.; Weinstein, C.E. (1994). *Student Motivation; Cognition and Learning. Essays in Honor of W. J. McKeachie.* (pp. 213-237). Hillsdale, N.J.: Lawrence Erlbaum Associates.

PIERS, E., & HARRIS, D.B. (1969). *The Piers-Harris Children´s Self-Concept Scale.* Nashville, TN: Counselor recordings and Tests.

PINTRICH, P. R. & DE GROOT, E.V. (1990 a). Motivational and Self-Regulated Learning Components of Clasroom Academic Performance. *Journal of Educational Psychology,* 82(1), 33-40.

POMERANTZ, S.C. (1979). Sex differences in the relative importance of self-esteem, physical self-satisfaction, and identity in predicting adolescent satisfaction. *Journal of Youth and Adolescence, 8* (1), 51-61.

POPE, A.W., McHALE, S.M., & CRAIGHEAD, W.E. (1988). *Self-esteem enhancement with children and adolescents.* Boston: Allyn & Bacon.

POZAR, F.F. (1983). *Inventario de Hábitos de Estudio.* TEA Ediciones S.A. Madrid.

ROSENBERG, M. (1979). *Conceiving the self.* Basic Books, New York.

SCHUNK, D. H. (1994). *Self-Regulation of Learning and Performance: Issues and Educational Applications.* Hillsdale, N.J.: Lawrence Erlbaum.

SELIGMAN, M.E.P. (1993). *What you can change and what you can´t.* New York: Fawcett Columbine.

SILVERSTEIN, L.R., STRIEGEL-MOORE, R.H., TIMKO, C., & RODIN, J. (1988). Behavioral and psychological implications of body dissatisfaction: Do men and women differ? *Sex roles, 19* (3/4), 219-232.

STREIN, W. (1988). Classroom-based elementary school affective education programs: A critical review. *Psychology in the Schools, 25,* 288-296.

URDAN, T.C. & MAHER, M.L. (1995). Beyond a Two Goal Theory of Motivation and Achievement. A Case for Social Goals. *Review of Educational Research,* 65(3), 213-245.

WEINER, B. (1985). An attributional Theory of Achievement Motivation and Emotion. *Psychological Review,* 92 (4), 548-573.

WEINER, B. (1990) History of Motivational Research in Education. *Journal of Educational Psychology,* 82 (4), 616-622.

WEXLER, D.B. (1991). *The adolescent self: Strategies for self-management, self-soothing, and self-esteem.* New York: Norton.

WISEMAN, C.V., GRAY, J.J., MOSIMANN, J.E., & AHRENS, A.H. (1992). Cultural expectations of thinness in women: An update. *International Journal of Eating Disorders, 11* (1), 85-89.

ZASTROW, C. (1994). Conceptualizing and changing the self from a rational therapy perspective. In T. M. Brinthaupt & R. P. Lipka (Eds.), *Changing the self* (pp. 120-175). Albany: State University of New York Press.

ZIMMERMAN, B. (1994). Dimensions of Academic Self-Regulation: A Conceptual Framework for Education. En Schunk, D.; Zimmerman, B. *Self-Regulation of Learning and Performance,* (pp. 3-21) Hillsdale, N.J.: Lawrence Erlbaum.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 147-173

LA FORMACIÓN INICIAL DEL PROFESOR DE EDUCACIÓN SECUNDARIA. SITUACIÓN ACTUAL Y PERSPECTIVAS DE FUTURO

Antonio Pantoja Vallejo
Tomás J. Campoy Aranda
Universidad de Jaén

RESUMEN

Hace bastante tiempo que la formación inicial del profesorado de Educación Secundaria está en entredicho. Si la aparición en los años 70 del Curso de Aptitud Pedagógica (CAP) supuso un avance en la concepción del acceso a la docencia, el paso de los años ha puesto de manifiesto el desfase y obsoletismo de este sistema y la incapacidad de las autoridades educativas para implementar nuevas formas de preparación de un profesorado que debe estar mejor preparado para hacer frente a un alumnado cada vez más diverso y multicultural. En el presente artículo, en el que tomamos como referencia las distintas normativas, analizamos toda esta problemática y, partiendo de una investigación llevada a cabo con alumnos del CAP, realizamos diversas propuestas dirigidas a los sectores implicados con el fin de contribuir a una mejora en la calidad de la preparación del profesorado de Educación Secundaria.

ABSTRACT

The initial training of Secondary Education teachers is in doubt for a long time. If the appearance of the Educational Aptitude Course (CAP) in the seventies meant an advance in the understanding of the access to the teaching, the passage of time has made evident the fact that this system is out of step with today's world and the inability of the education authorities to apply new forms of training to teachers who must be qualified to face up to very varied and

BEST COPY AVAILABLE

multicultural pupils. In this article, in which we take into account the state orders, we analyse all these problems. We start from an investigation that was carried out with the CAP pupils and we carry out some proposals to the involved sectors for helping to improve the quality of the teachers training of the Secondary Education.

1. LA FORMACIÓN INICIAL DEL PROFESORADO DE SECUNDARIA

La formación inicial del profesorado es considerada como uno de los medios más importantes e ineludibles para promover las reformas y mejoras que la sociedad demanda hoy de los sistemas educativos. Pero la búsqueda de la calidad en esa formación inicial debe pasar por una serie de premisas si no queremos quedarnos en unos planteamientos teóricos, que se van transmitiendo en algunas actividades puntuales, pero que no llegan a cuajar en propuestas concretas y viables para que puedan desarrollarse. Prueba de ello son los déficits provenientes de una escasa formación inicial muy patente en el profesorado de educación secundaria y que no tiene continuidad en la vida profesional de los docentes (Imbernón, 1994; Esteve, 1997; Campoy, 1997a, 1997b; Angulo, Etopa y Martín, 1998; Ferreres y otros, 1998; Yanes, 1998).

Durante años la formación inicial del profesorado de secundaria, que se ha venido realizando por medio del Curso de Aptitud Pedagógica (CAP), ha estado muy desprestigiada, cuando no sumida en un verdadero caos. En los últimos tiempos, algunas universidades han cerrado sus ICEs, y las Comunidades Autónomas con transferencias en educación han elaborado decretos de regulación del CAP en un intento de mejora de esa formación.

La realidad es que en el momento actual la mayoría de los profesores aprenden el ejercicio docente por intuición o mediante la práctica del ensayo y error. Todavía está muy presente que lo que un profesor necesita para dar clase es tener un buen dominio de la materia que va a impartir.

De forma general, la actual situación viene caracterizada por la ausencia de una cultura específica del profesorado de Secundaria, formado éste por un colectivo heterogéneo, sometido a una reconversión en gran medida impuesta. Todo esto repercute en el diseño de políticas de formación inicial y continuada del profesorado, en la que hay que atender a una enorme masa de personas, que tienen grados de implicación, experiencia y titulaciones muy diversas. Asimismo, la escasa motivación del alumnado del CAP, los problemas de planificación y coordinación del profesorado que lo imparte, los contenidos de formación, los horarios inadecuados, el elevado número de alumnos por aula, el excesivo contenido teórico de las clases, la inadecuación didáctica de las aulas, la falta de un profesorado estable y cualificado o la insatisfacción en la realización de las prácticas programadas, son algunas de las principales quejas manifestadas por el alumnado, que se traducen en un sentimiento generalizado de 'expectativas frustradas' y de 'trámite necesario' para el ejercicio profesional (Campoy, 1997a, 1997b; Esteve, 1997). Además de estos problemas por resolver quedan pendientes otros temas desde hace tiempo demandados, como son la reestructuración de los cuerpos docentes y sus competencias profesionales, el mantenimiento de la diploma-

tura en la titulación de Infantil y Primaria o la vieja reivindicación del cuerpo único de enseñantes.

Ante este panorama, se debe tener el valor suficiente de plantear la necesidad de la transformación de los planes de estudio de formación inicial del profesorado y de las instituciones y planes responsables de la formación permanente. Como señala Pérez Gómez (1990: 87) «ya es hora de integrar los programas de formación inicial, investigación educativa, innovación curricular y perfeccionamiento del profesor en proyectos convergentes».

El motivo de nuestro trabajo se encuentra, pues, en la inquietud que hemos mostrado como consecuencia de la experiencia de impartir módulos del CAP a lo largo de estos años, en la que hemos podido constatar, básicamente, los déficits importantes en los contenidos de los programas, la insuficiente formación inicial en su conjunto con que se ha pretendido formar al profesorado de secundaria, las discrepancias observadas y el grado de insatisfacción demostrado por los destinatarios de estos cursos de cualificación pedagógica.

Deseamos que nuestra aportación ayude a fomentar un proceso de reflexión no sólo en los profesionales de la docencia implicados en la formación inicial del profesorado, sino también —y lo más importante— en aquellas personas que han de tomar decisiones en la mejora de planes de estudios, de manera que se ponga en las mejores condiciones formativas al profesorado mediante la incorporación de los cambios que se están dando en el sistema educativo, al mismo tiempo que sea capaz de afrontar los restos de la sociedad actual.

2. LA EVOLUCIÓN EN LA FORMACIÓN DEL PROFESORADO DE SECUNDARIA

La filosofía de los programas de formación inicial del profesorado ha sufrido una considerable transformación en las últimas décadas, según se desprende del siguiente recorrido histórico.

Con la llegada de la II República se da un cambio en las perspectivas sobre la formación del profesorado. En efecto, el Plan de 1931 tiene un claro objetivo de profesionalizar la figura del maestro introduciendo una serie de cambios que afectará a su formación.

El desarrollo de este plan de formación del profesorado es paralelo a la aparición del primer plan sistemático de formación de profesores de segunda enseñanza en España. Es en 1932 cuando se crea a nivel oficial la Sección de Pedagogía, a la que corresponde la impartición de cursos para obtener el *certificado de aptitud pedagógica* dirigido a los profesores de enseñanza secundaria. Según Ruiz Berrio (1980) se dieron dos intentos previos para la institucionalización de la formación del profesorado de secundaria por medio de la Escuela Normal de Filosofía (1847-1852) y el Instituto-Escuela de Segunda Enseñanza de Madrid.

El franquismo tuvo como objetivo deshacer la obra educativa de la Segunda República y los ideales progresistas del siglo XIX. Durante esta etapa se configura un nuevo modelo escolar con un ideario de formación religiosa y patriótica.

El aperturismo de Ruiz Jiménez propicia la aparición de la Ley General de Educación de Villar Palasí (1970). Con esta ley «se culmina un proceso de implantación hegemónica de una visión técnico científica de la educación y de superación de la visión vocacionalista y patriotera de la profesión docente» (Yanes, 1998: 66). Se defendía la 'profesionalidad' docente mostrando la labor educativa como 'técnica' y el profesorado se alejaba de la reflexión y la búsqueda de alternativas en torno a los fines y las mismas condiciones bajo las que se impartía la enseñanza.

Esta ley, en su artículo 102.2b y la O.M. de 8 de junio (BOE de 12 de agosto), establece el modelo del CAP. La formación pedagógica será impartida por los ICEs y se desarrollará en dos ciclos. El primer ciclo será de carácter teórico y consistirá en el estudio de los fundamentos y principios generales de la educación necesarios para la labor docente. Las temáticas que se abordan son: a) Principios, objetivos y problemática de la educación en sus aspectos psicológicos, sociológicos e históricos; b) tecnologías y sistemas de innovación educativa; c) didácticas especiales. Se recomienda que haya un equilibrio entre la formación pedagógica general y las didácticas de las disciplinas correspondientes. Este primer ciclo tendrá como mínimo una duración de 150 horas, quedando al criterio de cada ICE las modalidades de realización. El segundo ciclo, de carácter práctico, consistirá en el ejercicio de la labor docente de los candidatos en los centros que se determinen, bajo la responsabilidad de los profesores tutores que se designen. El tiempo mínimo de este segundo ciclo será también de 150 horas.

La década de los ochenta se caracteriza por el inicio y progresiva extensión de una oferta de formación que intenta dar respuesta a la demanda generalizada en los sectores más comprometidos con la educación y su mejora. En este periodo se incrementa la oferta de formación, para adaptarse al nuevo marco curricular, institucional y profesional. Es una etapa de 'claroscuro' en la que se acentúan las contradicciones entre cantidad y calidad, extensión y profundidad que cuestionan las estrategias de formación y el sentido mismo de la formación al que éstas remiten: hay una errática política de formación permanente del profesorado, obsesivamente centrada en los aspectos cuantitativos de la formación y la ausencia de formación inicial profesionalizadora para gran parte del profesorado de Secundaria (Yanes, 1998).

Por su parte, la Junta de Andalucía por una Orden de la Consejería de Educación y Ciencia de 11 de diciembre de 1985 (BOJA nº 3 de 14-01-85) regula el curso para la obtención del CAP en las Universidades de la Comunidad Autónoma de Andalucía. Esta normativa señala que el proceso de formación inicial del profesorado es un elemento fundamental del sistema educativo y, con el fin de evitar desajustes en los objetivos que deben presidir el sistema de formación inicial, procede a reestructurar el Curso que tendrá en cuenta las siguientes normas básicas: a) Diseño curricular unificado; b) Diseño curricular integrador (teoría y práctica); c) Simultaneidad de las fases teórica y práctica; d) La fase práctica consistirá en la adscripción del alumno a un Seminario, participando en todas las actividades del mismo de una manera progresiva.

Los módulos que configuran el curso corresponden a Teoría de la Educación, Psicología de la Educación, Didáctica General y Didáctica Especial. La duración de

cursos —que simultánea la fase teórica y la fase práctica— es de 90 horas para la fase teórica y de 90 horas para la fase práctica. La evaluación del alumnado se basará en las entrevistas, control de asistencia y pruebas escritas, para la fase teórica. En la parte práctica se encargan los ICEs de la fase de simulación y los tutores de la fase de intervención.

La LOGSE establece en su artículo 24.2 que para impartir las enseñanzas en la etapa Secundaria es necesario, además, estar en posesión de un título profesional de especialización didáctica. Este título se obtiene mediante la realización de un curso de cualificación pedagógica, con una duración mínima de un año académico, que incluirá, en todo caso, un periodo de práctica docentes.

El Real Decreto de 20 de octubre de 1995 (BOE 9 de noviembre), completado con la Orden de 26 de abril de 1996 (BOE 11 de mayo), establece los contenidos y las características del nuevo Curso de Cualificación Pedagógica (CCP), con la finalidad de sustituir al criticado CAP. Este curso intenta proponer una solución, con algunas novedades interesantes —concretadas en otro apartado—, para ordenar y dar solidez a la formación inicial de los profesores de secundaria. Sin embargo, algunos elementos del modelo propuesto, y la inevitable inercia de las instituciones, puede convertir a este curso en una nueva versión del CAP, que herede sus limitaciones y su desprestigio (Esteve, 1997: 10). La novedad más importante de este decreto, es la consideración de las prácticas como elemento vertebrador del curso de cualificación pedagógica.

3. INVESTIGACIÓN SOBRE EL CAP

Nuestro punto de partida

Diversos autores han prestado recientemente su atención al CAP, como curso veterano en la formación inicial del profesorado de Educación Secundaria, por motivos distintos:

- Por sus lagunas en la formación del profesorado (Esteve, 1997).
- Por ser un terreno problemático y conflictivo (Escudero y otros, 1997).
- Para comprobar la idoneidad de un instrumento de análisis textual (Gil y otros, 1998)
- Para ofrecer propuestas de formación (Esteve, 1997; Yanes, 1998).

Incluso, algunas editoriales de implantación nacional han diseñado colecciones de libros destinadas a «rellenar» esos grandes vacíos formativos que la deficitaria formación de los profesores presentaba antes de iniciar su contacto real con los alumnos (ver, por ejemplo, la colección «Cuadernos de Formación del Profesorado de Secundaria» de la editorial ICE/Horsori).

Este inusitado interés por la formación del profesorado de Secundaria, unida a la falta de datos sobre los que establecer propuestas de futuro son los que nos han llevado a la realización de la investigación que ofrecemos a continuación de forma resumida.

Aspectos metodológicos

Objetivo

No es objetivo de esta investigación realizar inferencias en los resultados obtenidos en la misma, dado que no se controlaron variables ni se realizó un estudio poblacional. Nuestra intencionalidad ha sido tomar como referencia la opinión del alumnado que realizó el CAP durante el curso 98/99 en la Universidad de Jaén sobre aspectos diversos de tipo general, metodológico, contextual y organizativo en relación con los cuatro módulos que lo componen, con la finalidad de establecer líneas de trabajo futuras, que posibiliten un cambio y una adecuación en la formación inicial de los profesores de Educación Secundaria.

Por otro lado, es también objetivo del estudio reflexionar sobre las posturas personales ante los temas propuestos y ver hasta qué punto estos futuros profesores de Secundaria exponían algún tipo de soluciones a los problemas que ellos mismos planteaban.

Sujetos

La población la forman el total de estudiantes que realizaron el CAP en la Universidad de Jaén durante el curso 98/99, que asciende a 582 repartidos en 6 grupos. La muestra está compuesta por aquellos alumnos que voluntariamente decidieron expresar su opinión a través del cuestionario, cuyo número es de 383. A continuación, se expone la distribución muestral atendiendo a las principales variables de la misma. Aclaramos que algunos cuestionarios no incluyeron la titulación por lo que los totales de las titulaciones y ramas de estudio no concuerdan con el total de la muestra.

- Distribución por sexo:

	Frecuencia	Porcentaje	Porcent. válido	Porcent. acumulado
Hombre	170	22,8	44,4	44,4
Mujer	213	28,6	55,6	100,0
Total	383	51,5	100,0	

• Distribución por titulaciones:

Titulación	Frecuencia	Titulación	Frecuencia
Biología	19	Topografía	11
C. Superior de Música	9	Filosofía y Letras	2
Derecho	26	Farmacia	4
Enfermería	28	Ingeniería de Caminos	3
Estadística	2	Economía y Empresariales	33
Filología Francesa y Románica	1	Arquitectura	5
Filología Hispánica	26	Educación Física	1
Filología Inglesa	21	Psicología	19
Graduado Social	25	Geografía e Historia	8
Hostelería	1	Bellas Artes	1
Humanidades	10	Geodesia y Cartografía	2
Ingeniería Técnica	38	Sociología	2
Ingeniería Técnica en Informática	6	Ciencias de la Información	2
Pedagogía	1	Documentación	1
Química	18	Psicopedagogía	2
Relaciones Laborales	3	Matemáticas	3
Trabajo Social	23	Empresariales	8
Turismo	1	Física	2
Veterinaria	3	Fisioterapeuta	1

• Distribución por ramas de estudio:

	Frecuencia	Porcentaje	Porcent. válido	Porcent. acumulado
CC. de la Salud	35	5,9	9,7	9,7
CC. Experimentales	48	8,1	13,3	22,9
CC. Sociales y Jurídicas	112	18,8	30,9	53,9
Enseñanzas Técnicas	60	10,1	16,6	70,4
Humanidades y CC.EE.	89	14,9	24,6	95,0
Otros	18	3,0	5,0	100,0
Total	362	60,7	100,0	

Diseño y procedimiento

La investigación es de tipo descriptivo basada en la encuesta como forma de obtener la información. Dado el alto número de sujetos a los que se le aplicó el cuestionario, resultaba compleja la utilización de otras técnicas, quizás más directas y precisas como podría ser la entrevista personal. El cuestionario se pasó de forma simultánea en todos los módulos que componían el CAP en la Universidad de Jaén, para lo cual se solicitó al alumnado su cumplimentación confidencial y voluntaria y se les expuso la finalidad del mismo. Precisamente, la idea de la mejora de la calidad del CAP fue una constante durante el tipo de duración del curso y a ella se aludió en la presentación del cuestionario.

Instrumento utilizado

El instrumento utilizado ha sido un cuestionario de tipo mixto, que permitía al alumno valorar de forma cuantitativa cada uno de los ítems y expresar libremente su punto de vista acerca de los contenidos vertidos en los mismos. El cuestionario se compone de una parte de toma de datos personales, seguida de un total de 12 ítems cerrados que pedían al alumno una valoración entre 1 y 3 (de acuerdo, indiferente, en desacuerdo), completados con un ítem de valoración global del CAP (nada satisfactoria, satisfactoria, muy satisfactoria).

Los 12 ítems se refieren a las siguientes cuestiones:

1. Conocimiento que tiene el alumno de los objetivos y contenidos de las distintas materias.
2. Los profesores imparten clase de acuerdo con las características del grupo.
3. Metodología de aula adecuada a cada materia.
4. Materiales de trabajo ajustados a los contenidos de las materias.
5. Participación del alumnado en las clases.
6. El clima de aula ha favorecido las relaciones entre los alumnos.
7. Nivel de aprendizaje logrado de acuerdo con los objetivos planteados.
8. Modelos de evaluación seguidos.
9. Contenidos del CAP que preparan para el ejercicio de la docencia.
10. Bibliografía adaptada a los planteamientos de las distintas materias.
11. Labor desarrollada por el profesorado.
12. Contribución de los trabajos prácticos para la formación del futuro profesor.
13. Valoración global del CAP.

Finalmente, se incluyó un listado de aspectos importantes en la formación del profesorado, para que éstos manifestaran su interés sobre ellos o incluyeran otros que les parecieran importantes. Estos aspectos son: estrategias de aprendizaje, clima de aula, fracaso escolar, acción tutorial, técnicas de estudio, evaluación, legislación, organización del centro, atención a la diversidad, organización del currículum.

La fiabilidad del instrumento ha sido comprobada por varios estadísticos. La alfa de Cronbach muestra la alta consistencia interna del cuestionario con un índice de 0,7346. El modelo de las dos mitades ofrece una correlación todavía más alta (0,7752),

lo que demuestra que el planteamiento directo y la sencillez del cuestionario difícilmente pueden inducir a error en las interpretaciones.

Análisis de los resultados

Se ofrece a continuación un análisis conjunto de las respuestas cerradas y abiertas planteadas en el cuestionario. Este estudio de tipo comparativo nos permite un acercamiento más real al pensamiento del alumno.

Para realizar el análisis de las preguntas cerradas hemos utilizado el paquete estadístico SPSS para Windows en su versión 8. Para el estudio de los ítems de tipo cualitativo incluidos en el cuestionario se utilizó en un principio el programa de *Hyper Research*, no obstante, las primeras transcripciones de respuestas pusieron en evidencia la repetitividad de las mismas y la existencia de un corpus ya estratificado en función de frases cortas y directas en cada uno de los ítems. Esta circunstancia posibilitó que fuese bastante ágil la ordenación manual en función de las frecuencias de cada uno de los textos conforme éstos iban apareciendo.

- ***Ítem 1. Conocimiento que tiene el alumno de los objetivos y contenidos de las distintas materias:*** El nivel medio de grado de acuerdo con este ítem es de un 48,4%, siendo las mujeres las que manifiestan conocer mejor los objetivos y contenidos impartidos (50,5%). Es Ciencias de la Salud la rama que expresa un mayor acuerdo (67,7%).
  Comprobamos en las respuestas abiertas una unanimidad en el sentido de que no todas las materias han sido claras y explícitas para el alumno.

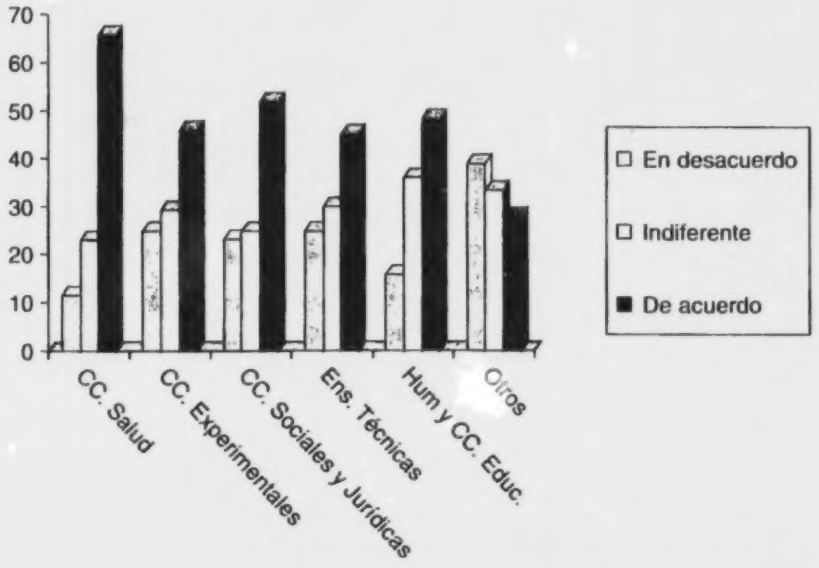


Gráfico 1
Valoración en relación con los objetivos y contenidos de las materias.

- *Ítem 2. Los profesores imparten clase de acuerdo con las características del grupo:* Un porcentaje por debajo de la mitad del alumnado encuestado (40,8%) expresa su acuerdo con la forma de impartir las clases del profesorado. No se aprecian diferencias significativas con relación al sexo. Son los alumnos de carreras pertenecientes a las ramas de Ciencias Sociales y Jurídicas y Enseñanzas Técnicas los que muestran un mayor grado de acuerdo.
  Del análisis cualitativo se desprende que la valoración depende de cada módulo, aunque en general se reconoce que el profesorado se ha esforzado en superar los inconvenientes propios del CAP. Un dificultad añadida ha sido la heterogeneidad de los grupos, debido a la diversidad de carreras y expectativas.

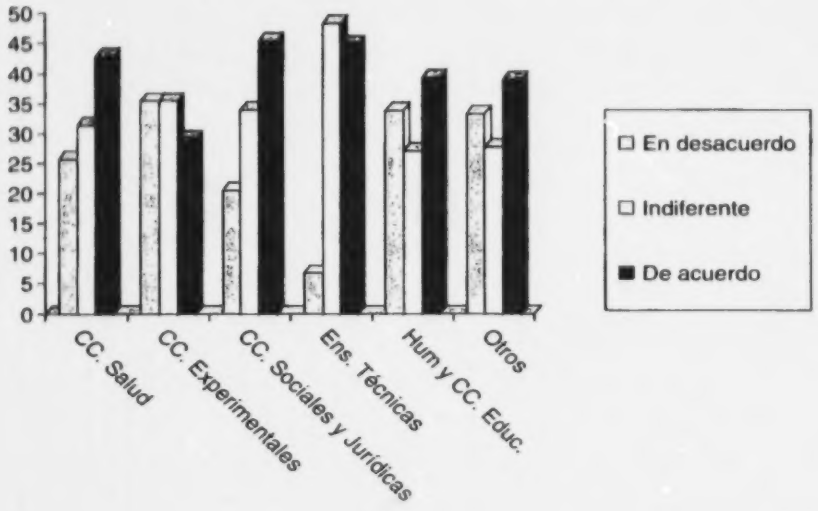


Gráfico 2
Adecuación de las clases a las características del grupo.

- *Ítem 3. Metodología de aula adecuada a cada materia:* Menos de un tercio del alumnado del CAP muestra su grado de acuerdo con la metodología empleada en cada materia. Destacan las mujeres con un 45,8% de desacuerdo, frente a un 31,8% de los hombres. Las ramas de Ciencias de la Salud y Ciencias Experimentales (35,4%) son las que reflejan un menor y un mayor grado de acuerdo, respectivamente.
  El análisis cualitativo pone de manifiesto el despiste del alumnado en este tema, a juzgar por el hecho de que no sepan bien qué se les está preguntando y no ajusten las respuestas a la pregunta planteada. No obstante, algunos de los encuestados manifiestan la dependencia de la metodología en función de la materia impartida.

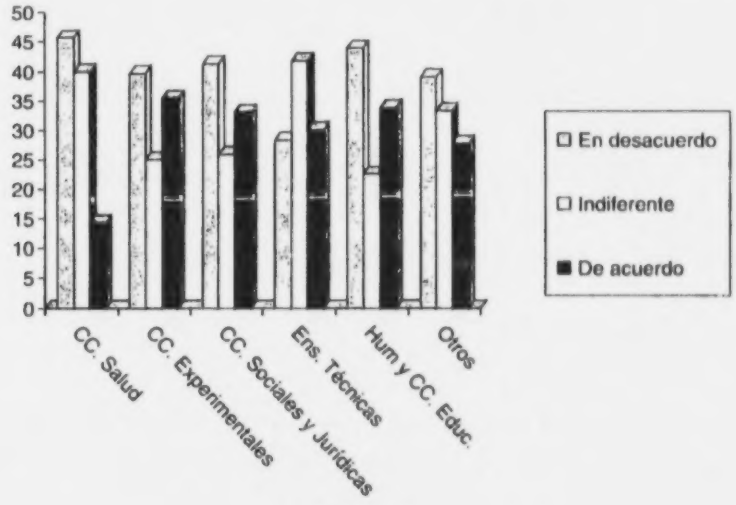


Gráfico 3
Metodología empleada.

- ***Ítem 4.*** *Materiales de trabajo ajustados a los contenidos de las materias:* Es el segundo ítem más valorado por los estudiantes (60,7%), en especial las mujeres con un 65,1%. Existe bastante igualdad en cuanto al grado de satisfacción entre las distintas ramas en relación con los materiales facilitados para su estudio y trabajo. Destaca Ciencias Experimentales con un 70,8%.
  En general el material de clase aportado ha sido valorado positivamente por el alumnado, tan sólo un grupo no significativo aboga porque se den menos apuntes y fotocopias.

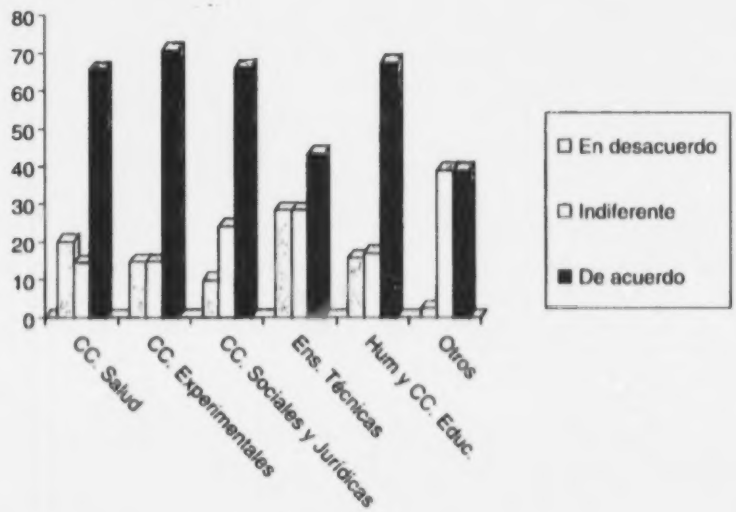


Gráfico 4
Relación entre materiales y contenidos de la materia.

- ***Ítem 5:*** *Participación del alumnado en las clases:* En general el alumnado se ha mostrado bastante satisfecho en cuanto a su participación en las clases, pues tan sólo el 16,8% muestra su disconformidad. La carrera que más muestra su grado de acuerdo (77,1%) es Ciencias de la Salud. A nivel global son las mujeres las que manifiestan un mayor grado de desacuerdo.
  En líneas generales ha sido positiva, aunque dependiendo de los módulos. Tanto la masificación por aula como el escaso tiempo del que se ha dispuesto, han sido factores que no ha favorecido la participación en el grado deseable. A pesar de todo, el ambiente ha sido distendido y colaborador.

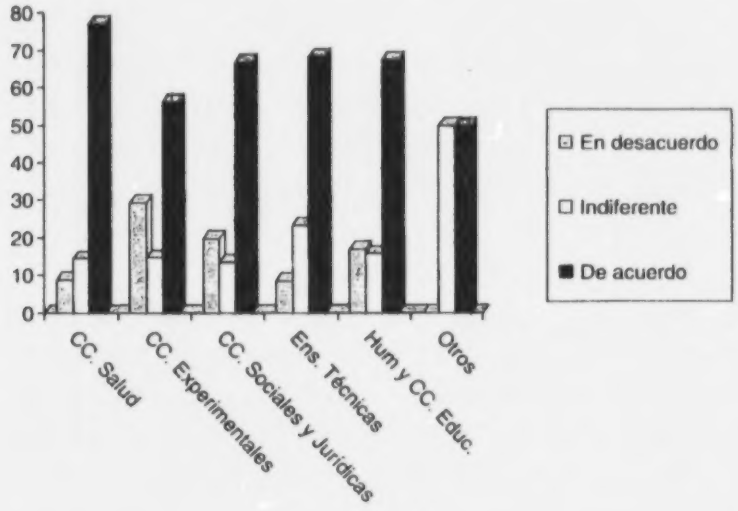


Gráfico 5
Participación del alumnado.

- ***Ítem 6.*** *El clima de aula ha favorecido las relaciones entre los alumnos:* El nivel de acuerdo con relación al clima de aula supera ligeramente a la mitad del alumnado (53,9%). No hay diferencia significativa por sexo. Las Enseñanzas Técnicas y Ciencias de la Salud expresan su mayor grado de satisfacción.
  De nuevo se observa la masificación como factor que no favorece las relaciones humanas. El tiempo de convivencia ha sido escaso, pero las relaciones son valoradas positivamente por el alumnado.

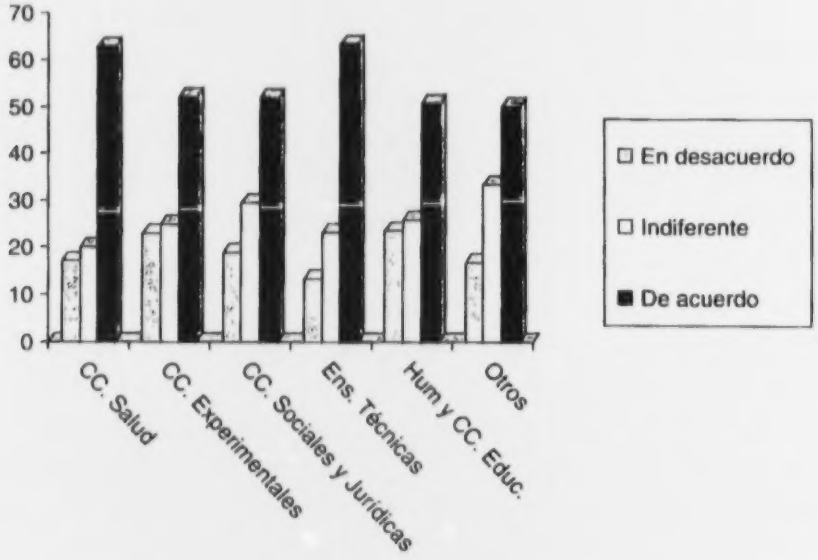


Gráfico 6
El clima de aula.

- ***Ítem 7.*** *Nivel de aprendizaje logrado de acuerdo con los objetivos planteados:* Algo menos de un tercio del alumnado del CAP (31,7%) se pronuncia favorablemente con relación al nivel de aprendizaje adquirido. No se aprecian diferencias significativas en cuanto al sexo ni a las ramas de estudios.
  La escasez de tiempo ha impedido que se profundizase suficientemente en algunas materias. Además, la condensación de los contenidos no les ha permitido absorver a un ritmo adecuado los contenidos.

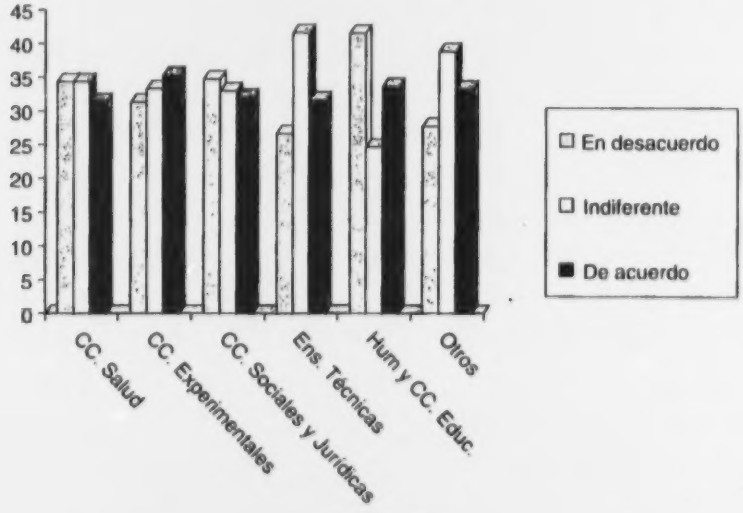


Gráfico 7
Nivel de aprendizaje alcanzado.

- ***Ítem 8.*** *Modelos de evaluación seguidos:* Tan sólo el 24,9% del alumnado se muestra satisfecho con los modelos de evaluación utilizados por el profesorado que imparte el CAP. El grado de desacuerdo es algo superior en las mujeres (53,8%) que en los hombres (48,2%). Ciencias Experimentales y Enseñanzas Técnicas tan sólo manifiestan su grado de acuerdo con el modelo de evaluación en torno al 18%.
  En general existe un rechazo al modelo tradicional de examen, inclinándose por una evaluación de tipo continua y formativa (trabajos, cuestionarios, etc.).

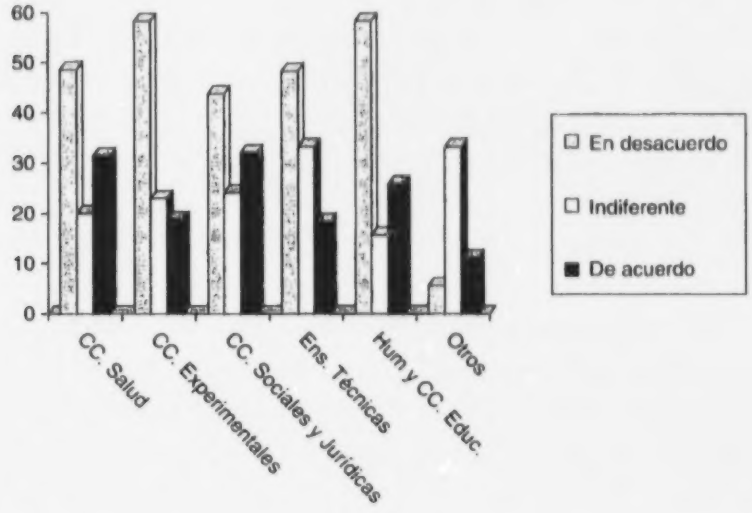


Gráfico 8
Modelos de evaluación.

- ***Ítem 9.*** *Contenidos del CAP que preparan para el ejercicio de la docencia:* El grado de acuerdo expresado por el alumnado con relación a los contenidos impartidos es un 23,6%. Las mujeres manifiestan un mayor grado de desacuerdo (54,2%), especialmente las tituladas de Enseñanzas Técnicas con un 68,8%. Existe unanimidad en las titulaciones en cuanto al bajo nivel de aceptación de los contenidos.
  Una parte del alumnado opina que aunque los contenidos son demasiados teóricos, pueden servir de base en su formación para la enseñanza. Su crítica se centra en la descoordinación entre la formación teórica y práctica, por lo que no se sienten capacitados para reflexionar sobre su preparación docente.

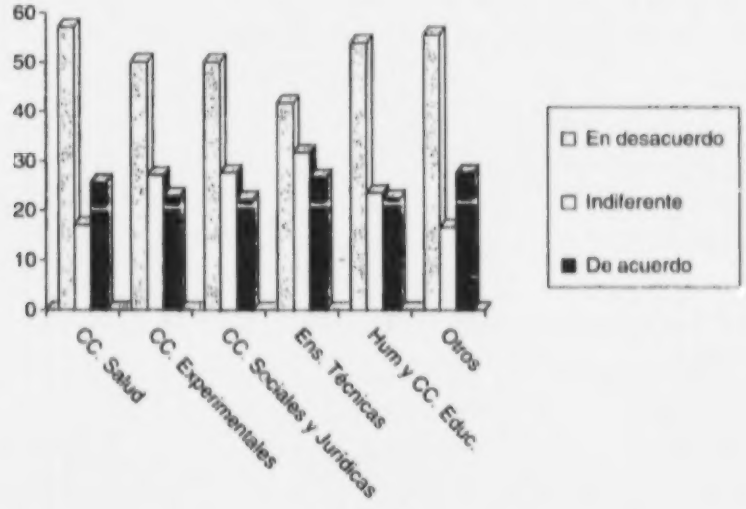


Gráfico 9
Contenidos del CAP.

- ***Ítem 10.*** *Bibliografía adaptada a los planteamientos de las distintas materias:* Un porcentaje inferior a la mitad (40,3%) está satisfecho con la bibliografía aportada por el profesorado a lo largo del curso. Los hombres están más en desacuerdo que las mujeres respecto al nivel de adaptación bibliográfica. Ciencias de la Salud y Ciencias Experimentales son las ramas que muestran un nivel más alto de acuerdo.
  No ha existido un criterio unánime a la hora de aportar la bibliografía al alumnado.

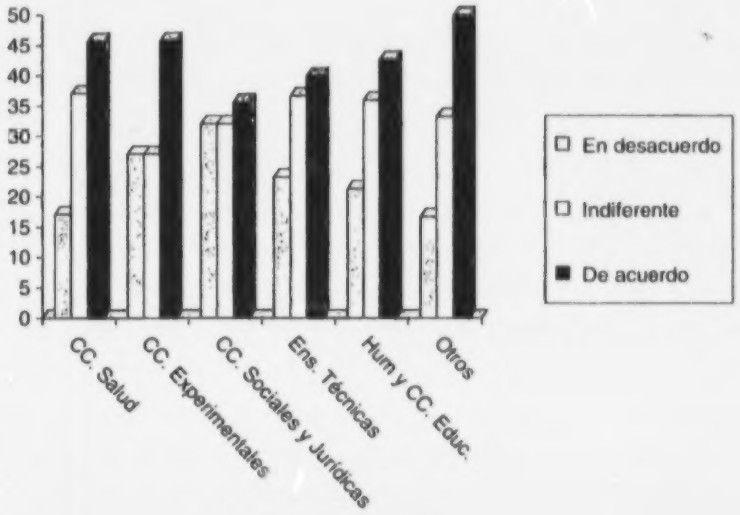


Gráfico 10
Adaptación bibliográfica.

- *Ítem 11: Labor desarrollada por el profesorado:* La mitad del alumnado considera que la labor del profesorado ha sido positiva (49,5%). Un porcentaje muy reducido se manifiesta en contra (15,7%). No se aprecian diferencias por sexos. Los titulados en Ciencias Sociales y Jurídicas son los que expresan una valoración más alta del profesorado.
  Se valora positivamente el esfuerzo del profesorado, a pesar de la masificación de las aulas (grupos de 100 alumnos) y de las limitaciones de tiempo.

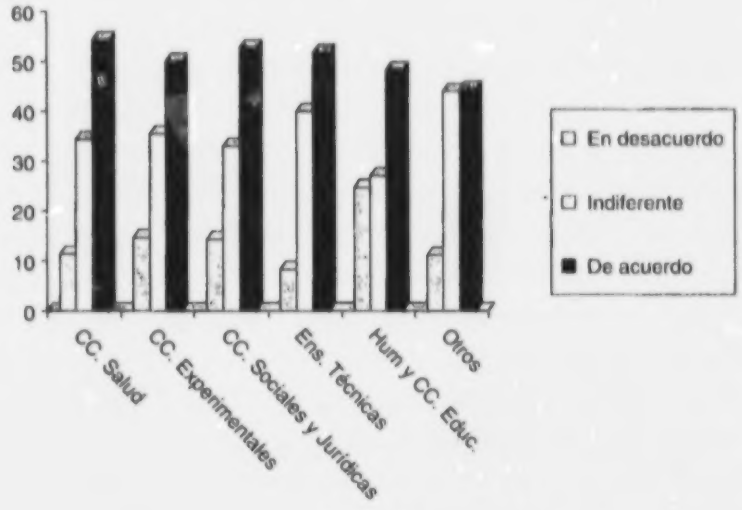


Gráfico 11
Labor desarrollada por el profesorado.

- *Ítem 12. Contribución de los trabajos prácticos para la formación del futuro profesor:* Un porcentaje inferior a la mitad del alumno (44,2%) se declara satisfecho con la realización de trabajos prácticos. No destaca ningún sexo en este ítem. La rama de Ciencias de la Salud muestra un alto índice de acuerdo (57,1%).
  El alumnado valora de forma positiva los cuestionarios utilizados en distintas materias como medio para interiorizar y reflexionar sobre cuestiones de interés.

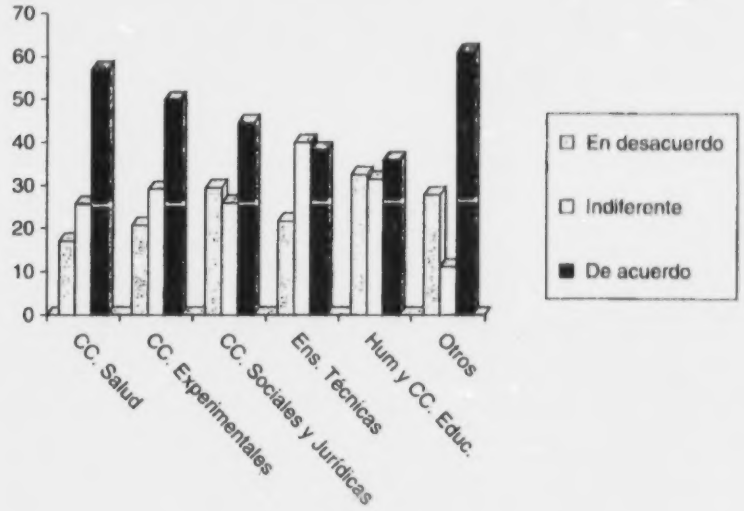


Gráfico 12
Trabajos prácticos.

- ***Ítem 13.*** *Valoración global del CAP:* Únicamente existe un 2,6% del alumnado que valora muy satisfactoriamente el Curso de Aptitud Pedagógica. No hay diferencias por sexos. La rama que muestra menos satisfacción hacia el CAP es Humanidades y Ciencias de la Educación (44,9%).
  En conjunto encontramos un amplio espectro de juicios de valor, la mayoría de ellos en una línea de claro reproche a la administración educativa responsable y al sistema educativo en general, cargados ambos de burocracia. De forma priorizada y ordenadas en positivas y negativas, se exponen a continuación las respuestas más repetidas:
  Positivas:
  a) Bien como informativo y preparatorio y como introducción al mundo de la enseñanza.
  b) Sería preferible un CAP más específico por materias y campos de estudio.
  c) Me ha servido para saber cómo se encuentra el sistema educativo.

  Negativas:
  a) Excesivamente teórico. Sería más interesante que existieran más prácticas.
  b) Insuficiente comparado con el nivel exigido para el ejercicio de la docencia.
  c) Poco tiempo para asimilar excesivos conceptos.
  d) Una desorganización y una pérdida de tiempo.
  e) Excesivos profesores, sobretodo en psicología.
  f) Poca relación entre las ideas expresadas y los efectos prácticos.
  g) No prepara para enfrentarse al trabajo de ser profesor.
  h) Excesiva masificación en las aulas.
  i) La asistencia obligatoria provoca rechazo.
  j) Existe solapamiento entre algunas asignaturas.

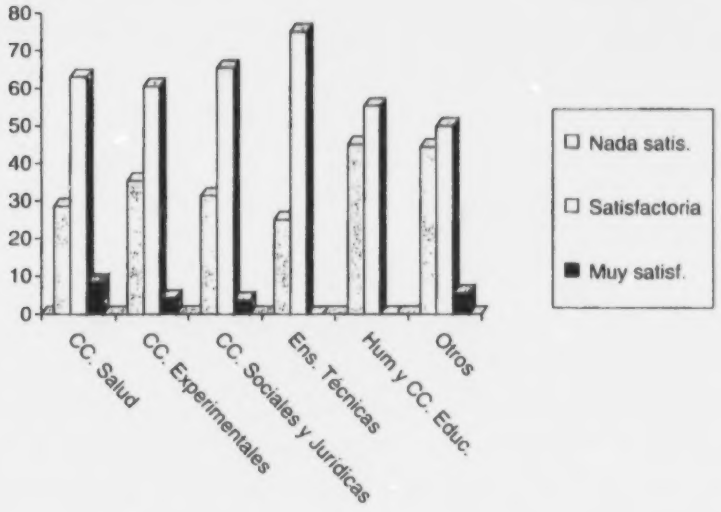


Gráfico 13
Valoración global del CAP.

El cuestionario incluía dos ítems de tipo cualitativo con la finalidad de recoger información a cerca de los motivos que llevaron al alumno a la realización del CAP y que priorizaran sus necesidades futuras de formación:

- *¿Cuáles fueron los motivos que te han llevado a la realización del CAP?* Esta pregunta tuvo como abanico de respuestas las que aparecen a continuación de forma priorizada:
  1. Para realizar oposiciones a Secundaria.
  2. Para completar mi formación.
  3. Abrir otros caminos o posibilidades laborales.
  4. Me gusta la enseñanza.
  5. Aumento del currículum.
  6. Para coger puntos en bolsas para formadores.
  7. Preparación para la docencia.
  8. Para conocer el mundo de la enseñanza.
  9. Reciclaje.
  10. Otros: curiosidad, cambiar de actitividad laboral, para organizar cursos, por entretenerme, por si me hacía falta, para ser un buen profesor,...
- *Necesidades futuras de formación.* La última cuestión incluida en el cuestionario nos ofrece la visión en orden a su importancia del profesorado con respecto a los temas considerados más importantes en relación a su formación:
  1. Estrategias de aprendizaje.
  2. Evaluación.
  3. Clima de aula.
  4. Atención a la diversidad.

5. Acción tutorial.
6. Técnicas de estudio.
7. Fracaso escolar.
8. Organización del currículum.
9. Organización de un centro de Secundaria.
10. Legislación sobre Educación Secundaria.

4. CONCLUSIONES

La primera conclusión, que además podría considerarse desconcertante, es que los profesores de Secundaria no realizan propuestas concretas ni ofrecen pautas de apoyo sobre las lagunas en su formación y las opciones que en un futuro se podrían abordar desde la administración.

El profesorado ha intentado superar los problemas inherentes a la propia estructura del CAP en un periodo de clara transición. Consecuencia de esto es que no siempre ha sido posible adecuar los contenidos previstos en la legislación a las nuevas necesidades de formación inicial. Esto ha podido provocar el que no se hayan trasmitido al alumnado de forma clara los objetivos y contenidos.

Los alumnos manifiestan que el grado de heterogeneidad de los grupos ha impedido una mayor calidad en la docencia impartida, al no haber sido posible atender a la diversidad formativa de las titulaciones de procedencia. El alumnado reconoce el esfuerzo realizado por el profesorado para paliar en lo posible esta situación.

El alumnado del CAP manifiesta su disconformidad con la metodología de aula que se viene utilizando. Esto nos hace pensar que la misma naturaleza de esta formación, distinta a la de una carrera universitaria, requiere el uso de metodologías más apropiadas a las finalidades propias del curso. Así por ejemplo, una necesidad manifestada por alumnos y profesores es trabajar con grupos menos numerosos que los actuales y disponer de aulas que permitan metodología más activas y participativas, así como otras formas de organización.

Los créditos asignados para el desarrollo de los distintos módulos no ha sido suficiente para cubrir las necesidades formativas. Asimismo, el carácter intensivo de la impartición del curso no ha propiciado un ritmo adecuado para la asimilación de los contenidos.

Existe un rechazo generalizado con relación al modelo tradicional de evaluación aplicado, lo que reafirma que no sólo se deben incorporar nuevas metodologías, sino formas renovadas de evaluación con un carácter más continuo y formativo.

El alumnado rechaza los contenidos impartidos en los distintos módulos por no contribuir a una formación que les permita afrontar la práctica docente en las aulas.

Con relación a la preparación para la docencia los alumnos manifiestan que no hay una articulación entre la teoría y la práctica, que se refleja en lo siguiente: por un lado, el aprendizaje adquirido no tiene un carácter efectivo, ya que no desemboca en el dominio de competencias más complejas indispensables para enfrentarse a las situaciones educativas reales; por otro, los contenidos de los diferentes módulos aparecen descontextualizados al no tener un reflejo en la fase de «practicum».

En síntesis, se puede afirmar que a nivel global sólo un pequeño porcentaje del alumnado (2,6%) se muestra muy satisfecho con las enseñanzas recibidas en el CAP, no obstante, a pesar de las deficiencias detectadas, el alumnado expresa su satisfacción por la labor desarrollada por el profesorado (84,3%), que ha intentando en todo momento superar las dificultades encontradas.

Existe una cultura muy extendida entre el alumnado de rechazo hacia el CAP, que se puede resumir en la siguiente frase, recogida de forma generalizada en los cuestionarios: «El CAP es un título que te piden, pero que no te capacita realmente para nada y que la gente no se toma en serio. Es un mero trámite».

Con relación a las necesidades formativas del alumnado del CAP, existe una tendencia hacia aquellas habilidades y competencias que le facultan para la práctica diaria en el aula, frente a otras actividades de tipo organizativo.

Del CAP al CCP

Todas las precariedades presentes en la concepción y desarrollo práctico del CAP puestas de manifiesto en nuestra investigación, fueron ya evidenciadas en la concepción misma de la LOGSE, que exigía exclusivamente la posesión de un título profesional de especialización didáctica para poder impartir clase en Educación Secundaria. El ya citado Real Decreto 1692/1995 de 20 de octubre de 1995 regula el denominado Curso de Cualificación Pedagógica (CCP), sustituto del CAP, y se completa con la Orden de 26 de abril de 1996.

En el preámbulo del Real Decreto queda clara la intención del CCP en el sentido de que «debe proporcionar al futuro profesorado de la educación secundaria la formación psicopedagógica y didáctica inicial necesaria para acometer las tareas propias de la función docente de acuerdo con los principios y fines establecidos en la LOGSE». Además se perfilan tres grandes consecuencias: de un lado la vinculación de estas enseñanzas a la universidad, de otro la vinculación de esta formación inicial con los institutos de secundaria y, por último, la relación entre la formación inicial de los profesores de formación profesional específica a los diversos sectores productivos.

Antes de entrar a analizar las virtudes y carencias de esta normativa, hay un detalle que, de entrada, llama la atención, y es que se obliga a las administraciones educativas a llevar a la práctica estas nuevas enseñanzas antes del curso 96/97, lo cual evidencia en la fecha que estamos que existe un cierto retraso en la aplicación de la ley —al menos en Andalucía—, lo que hace traslucir la complejidad que el mismo cambio en la concepción del CCP lleva implícita, tal y como veremos a continuación. En cualquier caso, el artículo 3 de la disposición adicional primera del citado real decreto, obliga a la generalización de estas enseñanzas a partir del curso 1999/2000.

La duración de estas enseñanzas de un curso académico o el hecho de que se organicen las clases en teóricas y prácticas son dos aspectos que no tienen nada de novedosos. Sin embargo al comprobar la distribución de los 60 a 75 créditos que dura el CCP, apreciamos diferencias claramente significativas y diferenciadoras de su antecesor CAP:

- Enseñanzas teórico-prácticas generales entre 19 y 23 créditos.
- Enseñanzas teórico-prácticas específicas entre 16 y 20 créditos.
- Enseñanzas teórico-prácticas optativas entre 5 y 7 créditos.
- «Prácticum» entre 20 y 25 créditos, de los que al menos 10 serán de docencia tutorizada.

Esta distribución de créditos no abarca a los profesores técnicos de formación profesional, que como novedad añadida del CCP, tienen un tratamiento diferenciado. Así, su carga lectiva estará entre 35 y 40 créditos, las enseñanzas generales entre 10 y 12 créditos, las específicas entre 8 y 10 créditos, las optativas entre 3 y 5 créditos y el «prácticum» entre 11 y 13 créditos, de los que 6 serán de docencia tutorizada. Aunque la normativa no lo aclara, pensamos que esta menor exigencia horaria es debida a que el volumen de materias que estos profesores impartirán en la Educación Secundaria es menor y de contenido especializado.

Resulta altamente significativa la distribución anterior por cuanto diferencia el currículum a impartir en tres grandes bloques, que posibilitan la adaptación a la naturaleza de cada carrera: general, específico y optativo. Precisamente, es esta concreción a cada situación personal, lo que nos parece más novedoso en la nueva normativa, por cuanto se propicia una formación diversificada en función de las carreras de origen y de acuerdo con las salidas profesionales de cada una de ellas.

Otro aspecto novedoso es la creación de la figura del profesor de Educación Secundaria asociado a la universidad, que se incorporará a los departamentos y participará en la impartición de las materias obligatorias específicas del bloque de enseñanzas teórico-prácticas y en la coordinación del «prácticum». Junto a este profesorado existirán los tutores del «prácticum», que atenderán a dos alumnos simultáneamente, emitiendo finalmente un informe en relación al trabajo desarrollado.

La masificación existente en las aulas del CAP se mejora en la normativa que venimos comentando reduciendo los grupos a un máximo de cincuenta alumnos.

La importancia dada a la parte práctica del CCP queda patente en la valoración global del «prácticum» del 50% de la puntuación final.

Por último, nos referiremos a las especialidades y a los núcleos formativos y de contenido que sustituyen a los antiguos módulos del CAP. Por un lado, se determinan un amplio número de especialidades para organizar el CCP, lo que favorece la adecuación de los contenidos a las distintas titulaciones de origen del alumnado. De otro se vertebran las materias obligatorias generales y las específicas. Entre las primeras destaca diseño y desarrollo del currículo, organización escolar, psicología del desarrollo y la educación, sociología de la educación y teoría e instituciones contemporáneas de educación, atención a la diversidad, alumnos con necesidades educativas especiales y tutoría y orientación educativa. Con esta distribución de contenidos se rellenan bastantes huecos existentes en la anterior configuración del CAP. En cuanto a las materias obligatorias específicas, podemos destacar las didácticas de cada área curricular y otras materias o módulos correspondientes a la especialidad del alumnado. Es curioso, que en ninguna de las normativas citadas aparezca mención alguna a las materias optativas, que como ha sido dicho tendrán una carga lectiva entre 5 y 7 créditos.

Situación actual

Resulta difícil señalar con claridad la situación en la que se encuentra la formación inicial del profesorado de Secundaria en estos momentos, sobretodo si pretendemos enmarcarla en una dirección distinta a la que ha venido siendo la habitual en los últimos años. La indiferencia de la administración educativa ha provocado que el Real Decreto 1692/1995 caduque antes de ponerlo en funcionamiento y exigir su cumplimiento a las CCAA con competencias educativas. Y es que decimos esto porque el MEC ha distribuido recientemente el borrador de un nuevo Real Decreto que modifica sustancialmente el anterior. Mientras tanto, se ha ido produciendo un cierto caos organizativo, de forma que algunas universidades como la de Barcelona o la de La Laguna han ofertado ya el nuevo CCP regulado en el Real Decreto que se va a modificar, mientras que otras —las más— han seguido impartiendo el obsoleto CAP, sin que exista una mínima autocrítica al sistema, ni se ofrezcan unas pautas de trabajo futuro en función de la larguísima experiencia acumulada con el paso de los años, ni se tenga en cuenta la opinión del profesorado que imparte esta docencia.

Nos referiremos, aunque sólo sea brevemente, a las novedades que ofrece este borrador, aun sabiendo que muchas de ellas pueden ser modificadas en su redacción definitiva.

Lo primero que destaca en este borrador es la intención de la administración educativa de retrasar la puesta en marcha del CCP, que no tendría carácter generalizado hasta el curso 2001/2002, perviviendo hasta esta fecha el antiguo CAP. Parece ser que existe una complejidad que escapa a nuestra perspectiva, que impide que las enseñanzas conducentes a la obtención del Título Profesional de Especialización Didáctica (TPED) entren en vigor de forma inmediata. Esto explicaría las prolongadas demoras.

En el preámbulo destaca la inquietud de la administración por «atender a los alumnos de diferentes culturas y étnias, alumnos con necesidades educativas especiales así como atender a la diversidad en el plano del diferente nivel de capacidad intelectual de los alumnos para que todos puedan avanzar aún más a pesar de sus distintas características personales e intelectuales...».

La nueva redacción del artículo 3 ofrece unos matices novedosos e, incluso, arriesgados. Es precisamente por considerar que el TPED se puede obtener por la vía normal de realización y superación del CCP y por una segunda vía, que incluso se coloca en primer lugar, que sería mediante asignaturas optativas y/o de libre configuración de los planes de estudios correspondientes. Desconocemos si estas asignaturas serán de nueva creación o mediante remodelación de algunas ya existentes. Matizamos ahora lo de arriesgado de este nueva perspectiva (apuntada ya por autores como Esteve, 1997), en el sentido de la problemática que encerraría llenar las clases de un alumnado variopinto, que lejos de enriquecer la multiculturalidad de las mismas puede crear no pocos problemas al profesorado para aunar posturas, intereses y motivaciones del alumnado específico de la titulación y los provenientes de las múltiples carreras que opten a la obtención del TPED por esta vía. Distinto sería —podría ser el caso— que la nueva perspectiva contemplase la creación de nuevas asignaturas

adaptadas al currículum contemplado en el anexo II del nuevo Real Decreto, que dicho sea de paso, encierra grandes novedades en lo que se refiere a las materias comunes de las enseñanzas teórico-prácticas. Principalmente destaca una mejor delimitación de los bloques y la inclusión de uno totalmente nuevo dedicado a Nuevas Tecnologías, lo que provoca que se pase de cuatro bloques a cinco, lo que lleva a pasar de dieciséis créditos a veinte. No obstante, se echa en falta un módulo dedicado a la iniciación en la investigación educativa y una mayor potenciación de la formación inicial del profesor en orientación y tutoría.

El artículo 4 en su apartado 3 parece que aclara el extremo anteriormente apuntado de la opcionalidad y la libre configuración y deja en manos de cada universidad la inclusión de las enseñanzas del CCP en los planes de estudio. Otro tema distinto es qué departamentos o áreas de conocimiento se hacen cargo de cada bloque de las enseñanzas teórico-prácticas.

Otra novedad incluida en el artículo 8 es la obligación del alumno de realizar un Proyecto de Fin de Curso, que suponga una mejora para el funcionamiento del Departamento Didáctico o para el centro en el que se realiza el CCP.

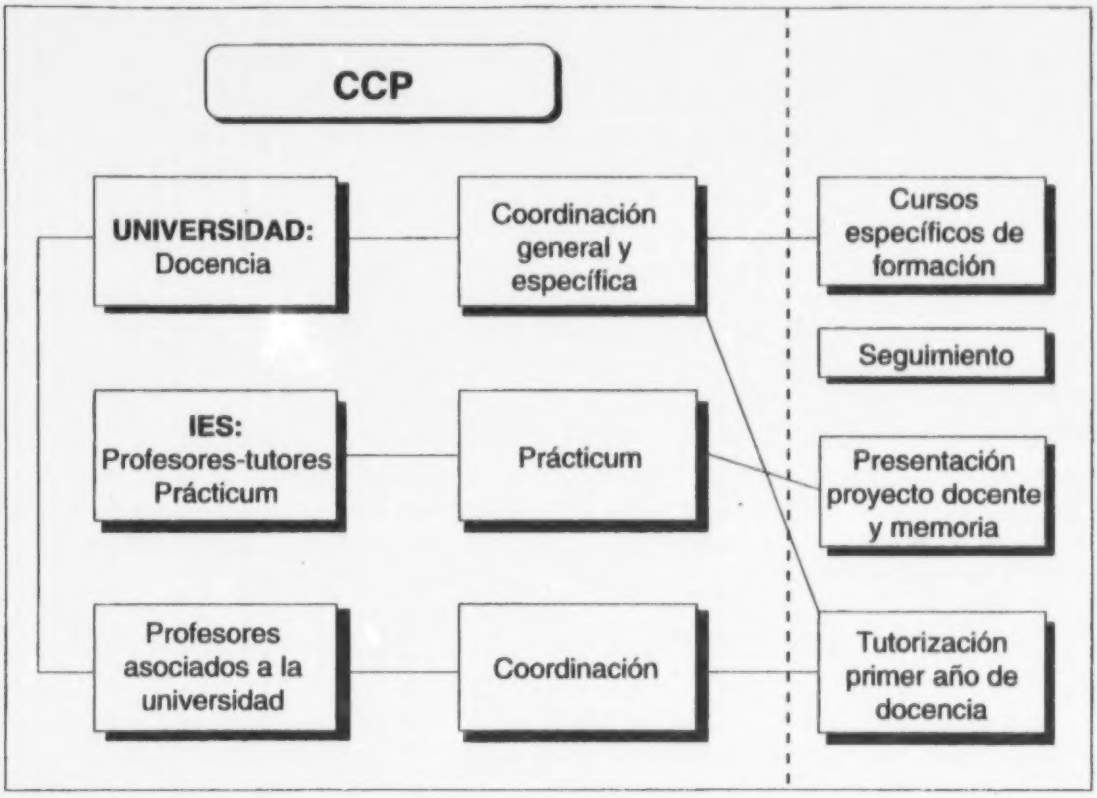
También cabe destacar en este apretado repaso del borrador una mejor concreción de las materias optativas, que echábamos en falta en el todavía vigente Real Decreto 1692/1995.

Por último, hay que señalar la reciente aparición del Real Decreto 321/2000 de 3 de marzo (BOE del 4), que modifica el Real Decreto 1692/1995. Este nuevo Decreto sólo recoge las propuestas de actualización del calendario de aplicación comentado anteriormente.

¿Cuál es el futuro de la formación inicial del profesorado de secundaria?

Nuestra postura personal pasa necesariamente por aunar las distintas perspectivas aportadas por la investigación, el contraste con la iniciativa institucional y nuestra propia experiencia como profesores del CAP. Existen claramente dos grandes opciones para obtener el TPED:

a) *Mediante asignaturas optativas o de libre configuración:* Sobre esta posibilidad hay que matizar la importancia que tiene que sean materias creadas por cada universidad expresamente para asumir esta función y que no deberán pertenecer a ninguna titulación, aunque sí se adscribirán a un área concreta de conocimiento.
b) La *realización del CCP* en un sentido similar al expresado en el borrador de Real Decreto, que sustituye parcialmente al Real Decreto 1692/1995. Esta organización queda reflejada en el siguiente gráfico.



A partir de la línea de puntos sugerimos la inclusión de una serie de aspectos que tendrían incidencia en ambas posiciones señaladas anteriormente, teniendo en cuenta que el «prácticum» sería idéntico para las dos vías de acceso al TPED. Comentaremos de forma sucinta cada uno de ellos:

- *Cursos específicos de formación* en los primeros años de docencia —especialmente en el primero— dirigidos de forma sectorizada y por especialidades, que además deberán contemplar un seguimiento de la proyección práctica de los conocimientos adquiridos. Nos inclinamos por un mayor número de cursos de fundamentación psicopedagógica frente a los propios de la especialización didáctica del profesor. Se trata de una responsabilidad que recae exclusivamente sobre la administración educativa.
- Realización de un *proyecto docente* sobre la especialidad del alumno y de una *memoria* de su actividad en el prácticum. Ambos instrumentos pueden situarse en un nivel más avanzado que el Proyecto de Fin de Curso contemplado en el borrador que comentamos anteriormente.
- *Tutorización del primer año de docencia* o realización de una fase práctica, que podría durar el primer año de trabajo una vez superada la fase de oposición. En aquellos casos que el alumno acceda a la Educación Secundaria mediante un contrato de interinidad también tendría que preverse un sistema de tutorización.

Concreción de nuestras propuestas

De las reflexiones procedentes de los anteriores apartados extraemos las siguientes propuestas que exponemos a continuación dirigidas a cada uno de los sectores implicados en la preparación multidisciplinar del profesor de Educación Secundaria:

- *A la Administración Educativa del Estado o de la Comunidad Autónoma:*
  - Coordinar las actuaciones de las distintas universidades.
  - Propiciar un cambio del modelo de oposición por otro más práctico y más adecuado a las nuevas necesidades del sistema educativo.
  - Continuar la tutorización del profesor una vez inicie su trabajo real durante un primer año de prácticas en los centros.
  - La previsión de cursos de formación continua que cubran el amplio abanico formativo del profesorado de Educación Secundaria, que lejos de terminar con la consecución del TPED, no hace más que empezar. Se trataría de que el profesorado fuese configurando su carrera profesional.
  - En sucesivas evaluaciones de las vías de acceso y de los contenidos que posibilitan el TPED, se podría considerar la ampliación del currículum con contenidos más cercanos a la relación con el alumno y la orientación académica y personal, que permitan una mejor preparación del profesorado para hacer frente a la tutoría de alumnos y de padres, así como en tareas orientadoras más adecuadas a las funciones del tutor recogidas en los distintos decretos.
  - Igualmente, consideramos imprescindible que la investigación educativa, materia básica en la formación de cada docente, se incluya en nuevas revisiones de los contenidos.
- *A la universidad, departamento o área de conocimiento:*
  - Unificar los criterios metodológicos de las distintas asignaturas con el fin de conseguir un cuerpo sólido y coherente, tal y como es característico de cada curso.
  - Realizar cursos de formación de formadores dentro de la misma universidad y seleccionar el profesorado que impartirá las asignaturas optativas o de libre configuración y del CCP.
  - Es imprescindible que exista un profesorado estable y con experiencia dedicado al CCP.
  - Conectar y evaluar de forma consecuente las clases teóricas-prácticas con el «prácticum», de forma que la evaluación basada en el modelo reflexión-acción proporcione información que revierta en la mejora permanente del modelo formativo.
  - Adscripción de cada materia a un área dentro de los departamentos de Pedagogía y Psicología de las universidades, lo cual permitiría procesos de autoevaluación y seguimiento de la docencia impartida.
  - Reconocer al profesorado la docencia impartida en las asignaturas de libre configuración y optativas y del CCP como enseñanza reglada.
  - Organizar unas jornadas de acogida en las que se informe al alumnado interesado sobre la forma de conseguir el TPED. Estas jornadas servirían de punto de

encuentro entre alumnos y profesores y entre profesores, tanto de universidad como de secundaria.

- *Al alumnado:*
  - Decidir al iniciar su formación en la titulación elegida su opción por la educación, con la finalidad de ir configurando su currículum sin prisas y tomar conciencia de las diferencias existentes entre el trabajo propio de su área de conocimiento y la educación de alumnos de Educación Secundaria.
  - Tomar conciencia intelectual y éticamente de la importancia de su formación inicial, tanto en cuanto trabajará para mejorar los contextos prácticos.
- *Al profesorado:*
  - Facilitar al alumnado antes del inicio del curso el programa con los objetivos y contenidos específicos de cada bloque o materia.
  - Establecer un tipo de metodología de aula de tipo crítico-reflexivo, que propicie la participación del alumnado.

Terminamos con una última aportación, que no es otra que la reivindicación de la orientación educativa como la llave que puede posibilitar que el profesorado que finalmente se dedique a la tarea de educar sea el más adecuado y preparado. En tal sentido, tendrá en un futuro cercano un peso específico —de hecho lo está teniendo ya en muchas universidades españolas— el Servicio de Orientación Educativa integrado en la misma estructura universitaria, que, entre otras funciones, podrá asesorar e informar a los recién llegados alumnos de las diversas titulaciones sobre la salida profesional en el ámbito educativo y sus peculiaridades intrínsecas.

BIBLIOGRAFÍA

ANGULO, Y., ETOPA, P., y MARTÍN, B. (1998). La formación inicial del profesorado. Sugerencias de mejora para la calidad desde la reflexión y el conocimiento de una realidad: las prácticas de enseñanza, *Congreso Nacional La Formación del Profesorado. Evaluación y Calidad*. Servicio de Publicaciones de la Universidad de Gran Canaria, 613-621.

CAMPOY, T.J. (1997a). Un estudio psicopedagógico del alumnado del CAP en la Universidad de Jaén, *La Orientación Educativa y la Intervención Psicopedagógica integradas en el Currículum*. AEOP: Valencia, 420-422.

— (1997b). Un análisis de necesidades en el alumnado del CAP de la Universidad de Jaén, *Actas del Congreso Nacional de Modelos de Investigación Educativa*. AIDIPE: Sevilla, 721-724.

ESCUDERO, J.M. y otros (1997). *Diseño y desarrollo del currículum en la Educación Secundaria*. Barcelona: ICE/Horsori.

ESTEVE, J.M. (1997). *La formación inicial de los profesores de secundaria*. Barcelona: Ariel.

FERRERES, V. (1998). Reflexiones y propuestas en torno a la formación del profesorado de Educación Secundaria, *Revista de Educación*, 317, 121-142.

GIL, J. y otros (1998). Análisis de respuestas libres en los cuestionarios. El método de las especificidades, *Revista de Investigación Educativa*, 14 (1), 129-147.

IMBERNÓN, F. (1994). La formación y el desarrollo profesional del profesorado. Graó: Barcelona.

LEY 14/1970, de 4 de agosto, *General de Educación y Financiamiento de la Reforma Educativa* (BOE del 6).

LEY ORGÁNICA 1/1990, de 3 de octubre, *Ordenación General del Sistema Educativo* (BOE del 4).

ORDEN DE LA CONSEJERÍA DE EDUCACIÓN Y CIENCIA, de 11 de diciembre de 1985 (BOJA nº 3 de 14-01-85).

ORDEN DE 26 DE ABRIL DE 1996 por la que se regula el *Plan de estudios y la impartición del Curso de Cualificación Pedagógica para la obtención del Título Profesional de Especializacion Didáctica* (BOE del 11 de mayo).

PÉREZ GÓMEZ, A. (1990). La formación del profesor y la reforma educativa, *Cuadernos de Pedagogía*, 181, 84-87.

REAL DECRETO 1692/1995, de 20 de octubre, por la que se regula el *Título Profesional de Especialización Didáctica* (BOE del 9).

REAL DECRETO 321/2000, de 3 de marzo, por el que se modifica el *Real Decreto 1692/1995, de 20 de octubre, que regula el Título Profesional de Especialización Didáctica* (BOE del 4).

RUIZ BERRIO, J. (1980). Estudio histórico de las instituciones para la formación de profesores. En *La investigación pedagógica y la formación de los profesores*. Granada: SEP, 99-120.

YANES, J. (1998). La formación del profesorado de secundaria: un espacio desolado, *Revista de Educación*, 317, 65-80.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 175-199

METODOLOGÍA PARA LA ENSEÑANZA DE LAS MATEMÁTICAS A TRAVÉS DE LA RESOLUCIÓN DE PROBLEMAS: UN ESTUDIO EVALUATIVO

Teresa González Ramírez
Métodos de Investigación y Diagnóstico en Educación.
Facultad de Ciencias de la Educación*
Universidad de Sevilla

RESUMEN

Este artículo expone los aspectos fundamentales tanto a nivel teórico como metodológico y las conclusiones más relevantes sobre la Evaluación de un Programa de Iniciación a las Matemáticas basado en la Resolución de Problemas para niños del Primer Ciclo de Educación Primaria. La Evaluación del Programa se ha realizado atendiendo a diferentes dimensiones evaluativas. Nos hemos centrado en la evaluación de los procesos de implementación del proceso resolutor a través del Esquema Lingüístico de Interacción (E.L.I.) previsto en el diseño del programa así como en los resultados o logros fundamentales que se han alcanzado durante su desarrollo. Para la evaluación del proceso resolutor hemos utilizado una escala de observación tipo lista de control; la evaluación de los resultados se ha realizado a partir de la elaboración de cinco pruebas de rendimiento teniendo como referentes evaluativos los objetivos del programa en las distintas áreas curriculares del mismo. Finalmente, esta investigación evaluativa, se aborda desde la percepción que sobre el programa han tenido los que lo han desarrollado considerando qué ha aportado a ellos el programa como docentes y a los alumnos que lo han recibido. Los resultados obtenidos señalan la aportación del programa como herramienta conceptual, la estimación de ciertos indicadores cualitativos de carácter actitudinal, organizativo y

* Av. San Francisco Javier, s/n. 41005- Sevilla. E-mail tgonzale@cica.es

social realizados por los profesores que lo han impartido así como un cambio de actitud ante la enseñanza de las matemáticas.

Palabras claves: *investigación evaluativa, educación matemática, resolución de problemas, educación primaria, lenguaje matemático.*

ABSTRACT

This article shows the fundamental aspects at both a theoretical and methodological level of and the most relevant conclusions concerning, the Evaluation of an Initiation Programme of Mathematics based upon problem solving for children at the first stage of Primary Education. The evaluation of this programme has been carried out taking into consideration different evaluative dimensions. We have concentrated on the evaluation of the process through the Interaction Linguistic Scheme (I.L.S) provided in the programme design as well as in the essential results and goals reached in its development. We have used an observation scale of the check list type for the evaluation of the resolving process. The evaluation of the results has been carried out from the elaboration of five performance tests taking as evaluative referents the objectives of the programme within the different curriculum areas of the very same programme. Finally, this evaluative investigation deals with the perception that those that developed it have over the programme, considering what the programme has given them us teachers and to the students who have received it. The results obtained signal the use of the programme as a conceptual tool, the estimation of certain qualitative indicators of attitudinal, organising, and social character carried out by the professors who have taught these classes as well as a change of attitude towards teaching mathematics.

Key words*: evaluative investigation, mathematics education, problem solving, primary education, mathematical language.*

MARCO TEÓRICO DE LA INVESTIGACIÓN

La enseñanza de las matemáticas constituye un campo de enorme interés científico. La sociedad actual, reclama el tener conocimientos matemáticos, resulta difícil encontrar parcelas del conocimiento en las que las matemáticas no hayan penetrado. Estudios realizados (Lapointe, Mead y Philips, 1989) muestran cómo la mayoría de las personas que no alcanzan el nivel de alfabetización mínimo como para desenvolverse en una sociedad moderna, encuentran las matemáticas aburridas y difíciles y se sienten inseguras a la hora de resolver problemas aritméticos sencillos; por otra parte, el tener conocimientos matemáticos se convierte en un importante filtro selectivo del sistema educativo. Esta situación plantea la necesidad de que, en la actualidad, los estudios en educación matemática se centren en la creación y desarrollo de estructuras didácticas formales que hagan más transparente el nexo de unión entre la cultura matemática experienciada por el niño antes del inicio de su escolaridad obligatoria y el conocimiento matemático de carácter formal que transmite la escuela. Los trabajos provenientes de la investigación transcultural, (Saxe, Guberman and Gearthart, 1993;

Rogoff and Lave, 1995) ponen de manifiesto como *la competencia matemática es un proceso cognitivo mediado por el tipo de actividad que desarrolla el sujeto (formal, no formal).*

De acuerdo a estos referentes, los tópicos que, en la actualidad son importantes desentrañar desde el status científico de la didáctica de las matemáticas son los que mostramos en la figura siguiente:

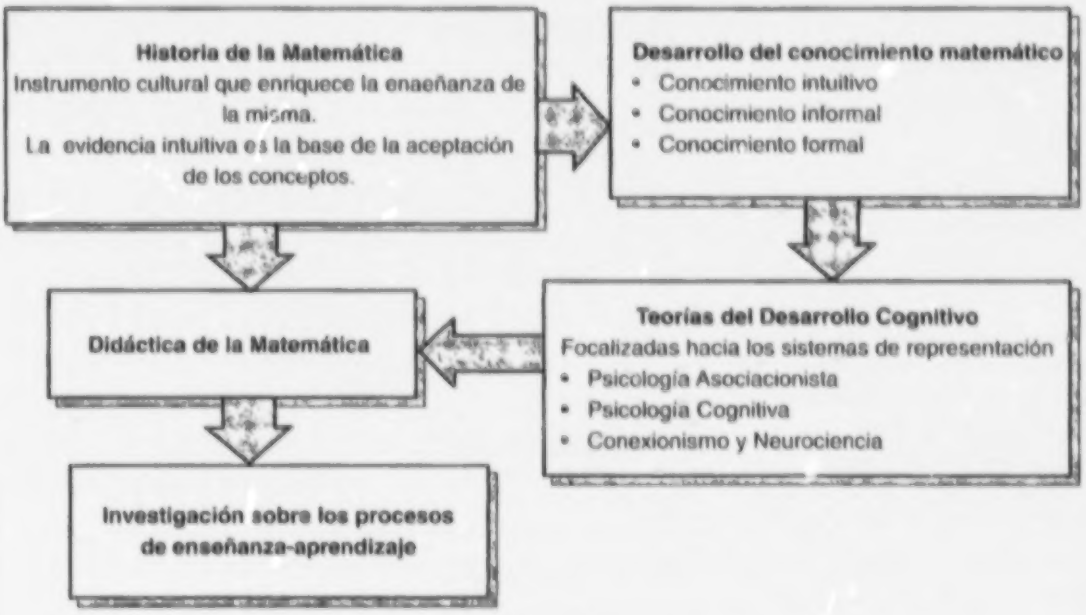


Figura 1

Referentes teóricos que confluyen en la educación matemática en la actualidad.

La comprensión completa y profunda de los conceptos fundamentales de una disciplina necesita del conocimiento de su historia, ya que ésta pone de manifiesto el proceso dinámico de la actividad científica como desarrollo permanentemente abierto, despertando en el sujeto que aprende unas actitudes y sobre todo unos hábitos metodológicos acordes con el método científico; de ahí de *la necesidad de partir de la evolución histórico-epistemológica de la matemática como ciencia* (González Urbaneja, 1991). La historia de la matemática es un instrumento cultural que enriquece la enseñanza de la misma. Este acercamiento a la historia de la matemática como ciencia nos dice que la evidencia intuitiva es la base de la aceptación de los conceptos, que las formulaciones de carácter formal e intuitivo preceden a la matemática exacta y formalizada y actúan como base para la misma; sin embargo, esta trayectoria no es la que ha marcado la forma de enseñarla. La enseñanza de la matemática está apoderada de una fuerte tendencia lógico-deductiva (Kline, 1978). Esa tendencia de carácter lógico-deductiva ha impregnado al conocimiento matemático de una serie de características (conocimiento formalizado, de naturaleza estrictamente abstracta, vinculado a un lenguaje muy específico y con propiedades que lo separan estructuralmente de los enfoques naturales, etc.) que no responde al carácter dinámico y evolutivo que la perspectiva

histórica señala y contradice resultados obtenidos en trabajos de investigación con relación al *desarrollo del conocimiento matemático*. Estos trabajos (Masingila, 1996; Lopes y Costa, 1996) consideran que, en muchos aspectos, el desarrollo matemático de los niños corre de forma paralela al desarrollo histórico de la matemática: el conocimiento impreciso y concreto de los niños se va haciendo cada vez mas preciso y abstracto. Según Baroody (1994) el proceso sucede de la siguiente forma: conocimiento intuitivo, conocimiento informal y conocimiento formal.

En la conjunción de estos dos elementos, *¿qué ha aportado la psicología como ciencia al estudio de cómo se construyen los procesos de representación en el niño?* La evolución paradigmática de la psicología como ciencia ha impregnado nuestro objeto de conocimiento de una serie de características (carácter sintáctico, semántico, lingüístico, etc.) que después han tenido un carácter prescriptivo sobre una teoría general del aprendizaje (en este caso matemático). Las aportaciones que, por ejemplo, desde la psicología cognitiva en sus distintas acepciones se han realizado polarizan por un lado, los planteamientos teóricos, que siguiendo a Vygotski postulan que el conocimiento tiene un origen eminentemente social y comunicativo y los teóricos que como Piaget sitúan el origen en el proceso mental interno del sujeto. Esta polarización establece grandes diferencias sobre una posible teoría del aprendizaje y en consecuencia establece diferencias substanciales al abordar didácticamente la enseñanza de las matemáticas. De ahí que desde los estudios que profundizan en la naturaleza de los procesos de representación en el niño sea importante construir una teoría del aprendizaje que nos permita crear estructuras formales que hagan más plausible el acercamiento al objeto de conocimiento. La *didáctica de las matemáticas* tiene actualmente el reto de profundizar en la creación y desarrollo de estructuras didácticas formales que hagan mas transparente el nexo de unión entre representaciones externas e internas (Maza, 1995). Las aportaciones realizadas sobre la evolución histórica de la matemática y el camino que habitualmente sigue el niño en la adquisición y aprendizaje de las nociones matemáticas fundamentales en conjunción con las aportaciones de la psicología de la educación, nos permite construir los fundamentos de una futura didáctica de las matemáticas.

Finalmente, esta presentación quedaría incompleta si no abordáramos qué aportaciones a nivel empírico se han realizado al campo de la enseñanza de las matemáticas. Conocer los ámbitos sobre los que se investiga es un reflejo de la evolución de esta disciplina. La revisión realizada al principio de los noventa por Putnam (1990) refleja los tópicos fundamentales sobre los que se investiga. Este autor plantea como fundamentales: *la psicología cognitiva*, centrada mas en el estudio de los procesos cognitivos y en la importancia de considerar la enseñanza de las matemáticas como una actividad generada en contextos de actividad práctica social y culturalmente organizadas; un segundo ámbito, centrado en el estudio de *las matemáticas como disciplina*, cuyo objeto es delimitar que actividades podrían ser consideradas matemáticas y finalmente, este autor señala el estudio de *las matemáticas como práctica docente*, centrada en las diferencias de conocimiento entre profesores expertos y noveles. A estos hay que añadir otra línea de investigación que en los últimos años ha despertado un gran interés, *la matemática como resolución de problemas*, desarrollando una serie de trabajos que pode-

mos englobar en una triple vertiente: a) Variables implicadas en la resolución de problemas, b) Estrategias en la resolución de problemas y c) Didáctica de la resolución de problemas. Una revisión actualizada realizada por Castro (1995) y Castro, Rico y Gil (1992) centrada en el estudio global o parcial de las tareas de resolución de problemas aritméticos plantea también estas corrientes investigadoras.

El esbozo de estos referentes teóricos constituyen los pilares básicos del Programa de Iniciación a las Matemáticas evaluado. La especificidad que tanto a nivel teórico como curricular adquieren en el diseño del mismo las presentaremos en el siguiente epígrafe.

BREVE DESCRIPCIÓN DEL PROGRAMA DE INICIACIÓN A LAS MATEMÁTICAS

En este apartado presentaremos una visión general del diseño y estructura del Programa de Iniciación a las Matemáticas (PIMRP) para alumnos del Primer Ciclo de Educación Primaria evaluado de forma experimental. Esta descripción se refiere, sobre todo, a los aspectos generales, los detalles más concretos pueden consultarse en González Ramírez (1997).

A nivel teórico se nutre de los pilares fundamentales de la Psicología Cognitiva (en sus distintas acepciones); en su génesis participa en algunos aspectos de la teoría del Procesamiento de la Información, sobre todo por lo que ésta aporta en contraposición al enfoque dualista de Descartes y la corriente introspeccionista (proveniente de la psicología freudiana) y asociacionista al estudio de los procesos psicológicos superiores.

Esta primera idea tomada como punto de partida hicieron que su autor (González del Pino, 1911-1986) se adentrara en las aportaciones que la psicología soviética, realiza al estudio de los procesos de representación en el niño; de ahí que aparezcan dentro de sus planteamientos conceptos claves como contexto problemático, problema, resolución de problemas, pensamiento matemático, esquema lingüístico de interacción, matizados y definidos todos ellos por los supuestos teóricos básicos del enfoque dialéctico-contextual.

Desde una perspectiva curricular el programa responde a una visión teóricamente coherente de cómo iniciarse en el aprendizaje de las matemáticas; se enmarca dentro del ámbito de la Didáctica de la Matemática que, a inicios de la década de los 80 vino a denominarse «Resolución de Problemas». *La finalidad básica del programa es conducir de acuerdo al perfil evolutivo del niño, con relación a la génesis y desarrollo del conocimiento matemático, la construcción del pensamiento matemático.* De acuerdo a esta finalidad el programa adopta una estructura curricular basada en la Resolución de Problemas que, a su vez, se constituye en la piedra angular del mismo siendo al mismo tiempo la herramienta que nos permite guiar el proceso de enseñanza-aprendizaje; intrínsecamente aporta una Metodología para la Didáctica de la Resolución de Problemas. Esta metodología se articula en torno al Esquema Lingüístico de Interacción (E.L.I.). El E.L.I., se convierte así, en un instrumento mediador basado en la interacción social dentro del aula y en el lenguaje, como herramienta conceptual que nos posibilita partir

del conocimiento experienciado por el alumno antes de iniciarse en la enseñanza obligatoria. Cada paso del Esquema Lingüístico de Interacción responde a un momento del proceso resolutor y nos permite conocer los procesos de representación que en el niño se están generando.

La Tabla I recoge cómo se desarrolla la interacción entre el profesor y el alumno durante el proceso resolutor y las distintas fases que componen la resolución de los problemas propuestos en el diseño del programa.

El desarrollo del programa se realiza en doce unidades de aprendizaje y cada una de ellas gira en torno a la consecución de unos objetivos generales y específicos sobre

TABLA I

LAS FASES DEL ESQUEMA LINGÜÍSTICO DE INTERACCIÓN. (E.L.I.)

FASE 1: COMPRENSIÓN DEL PROBLEMA
a) **Representación icónica del enunciado a través de un dibujo esquemático del mismo**. Debe responder a la pregunta: ¿qué sabemos de este problema? b) **Descripción verbal del enunciado dibujado.** Facilita esta descripción la representación mental del problema. El profesor pregunta: ¿qué queremos saber? c) **Aproximación a una estrategia de solución**. Para guiar ese proceso el profesor pregunta: ¿qué hay que hacer juntar, quitar o repartir?
FASE 2: EJECUCIÓN DE LA OPERACIÓN
a) **Supone la realización de la operación aritmética propiamente dicha.** La realización de la operación se realiza en dos momentos: Manipulando los elementos tangibles de la caja de cálculo (recursos enactivos) Haciendo una descripción verbal de los elementos que intervienen en la ejecución de la operación. El profesor pregunta: **¿Qué número se escribe primeramente?** (primer número de la operación a realizar) **¿Eso qué es?** (¿Qué representa el primer número de la operación a realizar?) **¿Qué número se escribe ahora?** (segundo número de la operación a realizar) **¿Eso qué es?** (qué representa el segundo número de la operación a realizar) **¿Qué se dice para hacer la operación?** (El alumno debe responder según el tipo de operación a realizar: juntar, quitar o repartir).
FASE 3: VERIFICACIÓN DE LOS RESULTADOS
Supone hacer un análisis de los resultados obtenidos. El profesor pregunta: ¿Ese número qué es? El alumno debe explicar qué significa el resultado.

la base de unos contenidos y actividades. Los contenidos del programa se estructura en relación a nueve áreas curriculares: Resolución de Problemas, Composición de Números Naturales, Las Horas, Composición de Números fraccionarios, Medición de Magnitudes, Geometría, Ángulos, Sistemas de Numeración, Operaciones y Cálculo.

Cada unidad de aprendizaje se inicia con la resolución de problemas (tres series de problemas con tres problemas cada serie). Desde el inicio del programa siempre se simultanean los problemas de juntar (sumar), quitar (restar) y repartir (dividir), seguidamente se realizan algunas de las actividades complementarias propuestas en el programa. Tanto la resolución de problemas como las actividades se estructuran en torno a un criterio de complejidad creciente. Las diferencias que se establecen entre las distintas unidades de aprendizaje en el área curricular «Resolución de Problemas» estriban en los siguientes aspectos: a) naturaleza de los problemas (gráfico, mixto y escrito), b) el nivel de representación de los problemas (posesión o pertenencia de un objeto/ valor o medida de objetos concretos), c) la numeración con la que se trabaja, d) la información que se ofrece en el enunciado del problema, e) el tipo de sentencia que se utiliza y f) la estructura semántica de los problemas.

A nivel metodológico resultan elementos claves del programa: atención más a los procesos que a los resultados, la importancia de la interacción verbal como base de la construcción del conocimiento matemático, el análisis y la reflexión metacognitiva sobre la resolución de problemas, estructura cíclica de las unidades de aprendizaje, el profesor como elemento mediador en el proceso resolutor y finalmente la utilización del Esquema Lingüístico de Interacción (E.L.I.) válido para todos los problemas del programa.

METODOLOGÍA DE LA EVALUACIÓN

Generalidades metodológicas del diseño evaluativo

Dentro de la investigación evaluativa, nuestro trabajo se centra en el ámbito de la evaluación de programas; ello supone atender a dos dimensiones fundamentales: una primera de carácter teórico-conceptual en la que resulta de vital importancia tener en cuenta los elementos básicos que en opinión de Colás y Rebollo (1993) delimitan el modelo de evaluación utilizado en la investigación. La tabla que presentamos a continuación, recoge las distintas dimensiones que debe contemplar un modelo de evaluación y las concreciones que adquieren en nuestro trabajo.

Desde una segunda dimensión de carácter metodológico el modelo evaluativo adquiere una serie de concreciones en cuanto a la delimitación de objetivos, hipótesis, técnicas de recogida de datos y de análisis de los datos obtenidos. Estas concreciones en nuestro trabajo siempre están focalizadas hacía los dos contenidos fundamentales a evaluar: el proceso resolutor y los resultados obtenidos por los alumnos durante la implementación del programa así como los logros que los profesores que lo han desarrollado le atribuyen en relación a lo que posibilita al alumno y a ellos como docentes.

TABLA II
EL MODELO DE EVALUACIÓN

EL MODELO EVALUATIVO	
CONCEPTO DE EVALUACIÓN	Valoración de los procesos y resultados de un programa
FINALIDAD DE LA EVALUACIÓN	Científica
CONTENIDOS A EVALUAR	Proceso/Producto
UNIDAD DE EVALUACIÓN	Sujetos (profesores y alumnos)
TOMA DE DECISIONES	Autoridad
PAPEL DEL EVALUADOR	Externo

Contextualización del estudio evaluativo

El estudio se desarrolló en dos centros públicos de Educación Primaria. En el primero de ellos contamos con los tres profesores de los tres grupos de primero y en el otro centro con los dos profesores de los grupos existentes para el nivel de segundo. En total fueron cinco profesores. El número de alumnos en cada grupo era bastante homogéneo. Para el nivel de primero, un grupo (G1) contaba con 19 alumnos y los otros dos (G2 y G3) con 21. Para el nivel de segundo un grupo (G4) estaba formado por 17 alumnos y el otro grupo (G5) por 15. En el curso de primero fueron un total de 61 alumnos y para el curso de segundo de 32.

La tabla adjunta resume la distribución definitiva según el centro (CEN1/CEN2) y el nivel educativo (N1/N2):

TABLA III
SELECCIÓN MUESTRAL SEGÚN NIVEL EDUCATIVO Y CENTRO

NIVEL DE PRIMERO/CENTRO 1		NIVEL DE SEGUNDO/CENTRO 2	
N° DE GRUPOS	ALUMNOS	N° DE GRUPOS	ALUMNOS
G1	19	G4	17
G2	21	G5	15
G3	21		

Selección de los participantes

Dado el carácter experimental del programa, la selección de los profesores que participaron en el estudio se realizó en conjunción con el Centro de Profesores de Sevilla y los responsables de Educación Primaria y del Área de Matemáticas a partir de un curso de formación en esta metodología que impartimos a todos los profesores

que durante ese curso escolar estuvieran en el primer ciclo de Educación Primaria. El desarrollo del curso pretendía un doble objetivo:

a) Retomar una labor iniciada por el autor del programa basada en la formación del profesorado en esta metodología desde un acercamiento a su propia práctica.
b) Iniciada esa labor formativa invitar a los profesores a participar en el estudio evaluativo.

El Curso de Formación se desarrolló en tres fases: de iniciación, seguimiento y evaluación. La **Fase de Iniciación** comenzó con una exposición sobre las bases conceptuales del programa. La **Fase de Seguimiento** se realizó con los tres profesores del nivel de primero de forma rigurosa y exhaustiva durante todo el proceso de implementación del programa. En estos profesores nos hemos basado para aportar toda la información relativa a la evaluación del proceso resolutor. La evaluación de los resultados del programa la hemos realizado con los profesores del nivel de primero y de segundo.

Respecto al grupo de profesores de segundo, tomamos las siguientes decisiones: a) Mantener la misma estructura didáctica que en el de primero, a diferencia de aumentar la numeración con la que se trabaja en los problemas planteados y b) Trabajar todas las actividades propuestas en el programa. Hay actividades de las que se contemplan en el programa (números fraccionarios, geometría, ángulos, etc.) que, a juicio de los profesores, exceden los objetivos de este primer ciclo de Educación Primaria según las directrices generales de los programas actuales para ese ciclo educativo.

La **fase de Evaluación** se organizó en torno a tres sesiones de evaluación que se realizaron a lo largo del curso con los cinco profesores. Cada una de ellas respondía a un objetivo evaluativo sobre el programa; el tipo de información aportada en estas sesiones de evaluación han apoyado la evaluación del programa desde la óptica del profesorado que lo ha llevado a la práctica.

2. EVALUACIÓN DEL PROCESO DE RESOLUCIÓN DE PROBLEMAS

La evaluación del proceso resolutor se plantea como objetivo fundamental conocer el grado de asimilación o incorporación en los alumnos del procedimiento resolutor (integración del Esquema Lingüístico de Interacción).

Este objetivo general se concretará en los siguientes objetivos más específicos:

a) Conocer si el alumno diferencia la estrategia de resolución de problemas: comprensión del problema, ejecución de una estrategia de solución y verificación o análisis del resultado obtenido.
b) Conocer la integración que el alumno ha realizado del Esquema de Interacción que le ayuda a diferenciar las tres partes del proceso de resolución de problemas (qué sabemos de este problema, qué queremos saber, qué hay que hacer juntar, quitar o repartir; etc...) y si existe alguna relación entre la integración que realiza el profesor y la que realiza el alumno.
c) Conocer que comprensión del problema, ejecución de la operación, verificación del resultado obtenido y grado de resolución total tienen los sujetos durante el proceso resolutor.

d) Conocer si es la caja de cálculo una herramienta adecuada para facilitar el paso hacía el razonamiento abstracto desligado de la manipulación de elementos tangibles.
e) Conocer si existen diferencias entre los sujetos en cuanto a comprensión, ejecución, verificación y grado de resolución total durante el proceso resolutor atendiendo a la naturaleza del problema (gráfico, mixto o totalmente escrito).
f) Conocer si existen diferencias entre los sujetos en comprensión, ejecución, verificación y grado de resolución total durante el proceso resolutor según el profesor que ha implementado el programa.
g) Conocer si existen diferencias entre los sujetos en cuanto a comprensión, ejecución y verificación durante el proceso resolutor según el nivel de resolución (alto, medio y bajo) de los sujetos.

Con la finalidad de alcanzar los objetivos señalados se planteó una recogida de datos basada en la observación del proceso resolutor durante todo el curso escolar. La lista de rasgos a observar es la siguiente:

LISTA DE RASGOS A OBSERVAR

h) ¿Realiza un dibujo esquemático del problema?
b) ¿Qué sabemos de este problema?
c) ¿Qué queremos saber?
d) ¿Qué hay que hacer juntar, quitar o repartir?
e) ¿Utiliza la caja de cálculo como material básico para la realización de la operación aritmética?
f) ¿Qué número se escribe primeramente?
g) ¿ Eso qué es? (¿Referido al primer número de la operación a realizar?
h) ¿Qué número se escribe ahora?
i) ¿Eso qué es? (¿Referido al segundo número de la operación a realizar?
j) ¿Qué se dice para hacer la operación?
k) ¿Ese número qué es? (¿Referido al número que se ha obtenido como resultado en el problema?

Estas preguntas que conducen el proceso resolutor en situación de interacción (profesor-alumno) se han utilizado tanto para conocer el grado de fidelidad con que reproducen los profesores la secuencia del proceso resolutor así cómo la incorporación del mismo que realiza el alumno; de ahí que la información recogida sea tanto del profesor como del alumno.

El listado de rasgos incluye las tres etapas fundamentales del proceso resolutor: comprensión del problema, ejecución de una estrategia de solución y verificación o análisis del resultado obtenido al concluir la operación matemática.

En total se han realizado doce sesiones de observación de las cuales las cuatro primeras tienen un carácter exploratorio con objeto de indagar sobre la viabilidad de la lista de control elaborada de acuerdo con el contexto de observación y los objetivos

del estudio. El criterio que se ha utilizado para la selección de estos períodos de tiempo intersesionales ha sido el que cada uno de ellos fuera representativo de las distintas unidades de aprendizaje que se realizan a lo largo del programa, de tal forma que el conjunto de todas las sesiones de observación supone una muestra de la variabilidad que implica por parte del alumno y el profesor la resolución de problemas en el programa. Cada sesión se ha dividido en números de intervalos de longitud constante y en el transcurso de la misma se anota si el proceso resolutor se conduce de la forma prevista. Es un registro de todo o nada. Nos interesaba conocer cómo cada profesor guiaba el proceso resolutor a partir del Esquema de Interacción propuesto y cómo respondían los alumnos seleccionados a cada una de las preguntas que el profesor hacía en relación al proceso resolutor. Para optimizar los registros realizamos cuatro sesiones de observación que tenían un carácter exploratorio con objeto de controlar la posible reactividad de los sujetos; estas sesiones también se utilizaron para mejorar la lista de rasgos a observar.

La tabla que presentamos a continuación recoge el total de observaciones realizadas a lo largo del curso:

TABLA IV
NÚMERO DE REGISTROS OBSERVACIONALES REALIZADOS

OBSERVACIÓN DEL PROCESO RESOLUTOR								
Profesor 1			Profesor 2			Profesor 3		
Sujeto 1	Sujeto 2	Sujeto 3	Sujeto 4	Sujeto 5	Sujeto 6	Sujeto 7	Sujeto 8	Sujeto 9
Observación 1 ..... Observación 8			Observación 1 ..... Observación 8			Observación 1 ..... Observación 8		

Para la selección de los sujetos nos basamos en los trabajos realizados por Brown (1983) y Carpenter y Moser (1984). Estos autores señalan como componentes esenciales en la resolución de problemas matemáticos procesos de carácter metacognitivo y afectivo, destacando la importancia de fomentar una reflexión sobre la práctica en el alumno y una motivación basada en la propia tarea.

Los factores que observamos en el aula durante el desarrollo del programa fueron los siguientes: a) planificación de la tarea de cara al entendimiento de la misma, b) control de actividades durante la propia resolución, c) tipo de representación efectuada sobre el problema propuesto y d) nivel de motivación hacía la tarea. Cada sesión de observación estaba focalizada hacía cada uno de esos sujetos durante un intervalo de tiempo corto y por turnos; dividimos el período de observación en tantos intervalos como individuos y dedicamos a cada uno de ellos un intervalo de tiempo.

La tabla siguiente resume todo el proceso metodológico desarrollado para la evaluación del proceso resolutor:

TABLA V
CUADRO-RESUMEN SOBRE METODOLOGÍA UTILIZADA PARA LA EVALUACIÓN DEL PROCESO RESOLUTOR

LA EVALUACIÓN DEL PROCESO RESOLUTOR	
OBJETIVOS EVALUATIVOS	a) Nivel de implementación del programa con relación al proceso resolutor a través de la interacción profesor-alumno durante el proceso resolutor y las diferencias que se establecen entre los sujetos según el profesor, el tipo de alumno y los niveles de resolución de los alumnos.
INSTRUMENTOS	Observación sistemática. Escala de observación tipo lista de control (las preguntas del proceso resolutor constituyen la lista de rasgos a observar).
TÉCNICAS DE ANÁLISIS	a) Análisis descriptivo basado en las once preguntas-respuestas sobre las que se analiza el proceso resolutor. b) Análisis de la varianza para establecer si existen diferencias en el proceso resolutor según el profesor que lo ha desarrollado, la naturaleza del problema, niveles de resolución de los sujetos y por la interacción profesor y niveles de resolución de los sujetos.

LA EVALUACIÓN DE LOS RESULTADOS DEL PROGRAMA

La evaluación de resultados tiene como objetivo general conocer cuáles han sido los principales logros del programa. Este objetivo general, se canaliza a través de dos agentes: profesores y alumnos.

En relación al alumno este objetivo general se concreta en los siguientes objetivos mas específicos:

a) Conocer el grado de consecución de los objetivos del programa (en sus distintas áreas curriculares) en los alumnos del nivel de primero y segundo de Educación Primaria con los que se ha desarrollado el programa.

De este objetivo prioritario para esta evaluación del programa de deriva otro objetivo que en nuestro estudio tiene un carácter más secundario:

b) Conocer si existen diferencias entre los alumnos de los distintos grupos (tanto del nivel de primero como de segundo) en cuanto al grado de consecución de los objetivos del programa según el profesor que lo ha llevado a la práctica.

En relación al profesor nos hemos planteado los siguientes objetivos más específicos:

c) Conocer cómo valoran los profesores los logros alcanzados en el programa en cuanto a la consecución de objetivos, contenidos, actividades, recursos y aspectos organizativos del mismo.
d) Conocer qué dimensiones del programa valoran los profesores en relación a cómo ha incidido el desarrollo del programa en los alumnos.
e) Conocer qué dimensiones del programa valoran los profesores en relación a la influencia que ha ejercido en ellos durante su desarrollo.

La tabla siguiente resume todo el proceso metodológico desarrollado para la evaluación de los resultados del programa:

TABLA VI
CUADRO-RESUMEN SOBRE METODOLOGÍA UTILIZADA PARA LA EVALUACIÓN DE LOS RESULTADOS DEL PROGRAMA

LA EVALUACIÓN DE LOS RESULTADOS DEL PROGRAMA		
	SOBRE EL ALUMNO	**DESDE EL PROFESOR**
OBJETIVOS EVALUATIVOS	a) Conocer los logros alcanzados en el programa a través del nivel de consecución de objetivos en las distintas áreas curriculares.	a) Valoraciones que hace el profesorado en cuanto a consecución de objetivos, contenidos, etc... del programa. b) Valoraciones del profesorado en cuanto a cómo ha incidido el desarrollo del programa en los alumnos y en ellos como docentes.
INSTRUMENTOS	Pruebas de carácter criterial para medir los objetivos del programa en el nivel de primero y de segundo	Sesiones de evaluación guiadas por entrevistas semiestructuradas.
TÉCNICAS DE ANÁLISIS	a) Análisis descriptivo para comprobar el nivel de consecución de los objetivos del programa. b) Análisis de la varianza para ver si existen diferencias entre los logros alcanzados por cada grupo y nivel educativo.	Análisis de contenido: se toma como unidad de análisis la unidad temática conceptual significativa.

En relación a la **evaluación de los resultados del programa**, dos han sido las fuentes principales de obtención de los datos:

Cinco **Pruebas** tanto para el nivel de primero como de segundo, para evaluar el rendimiento matemático de los alumnos teniendo como referentes o criterios de comparación los objetivos del programa.

Tres **Sesiones de Evaluación** realizadas con los profesores durante la implementación del programa con objeto de analizar e indagar qué dimensiones del programa resultan, a juicio de los profesores, más reveladoras de la posible efectividad del mismo; incidiendo en este análisis en la influencia que haya podido ejercer tanto en los alumnos como en los profesores. Los objetivos que pretendían cubrir cada una de ellas son los siguientes:

Objetivos de la Primera Sesión de Evaluación

- Indagar en la trayectoria personal de cada uno de los profesores en cuanto a sus conocimientos, expectativas y deseos en relación a la Didáctica de las Matemáticas.
- Explorar las primeras impresiones sobre el programa tras una primera toma de contacto con él.

Objetivos de la Segunda Sesión de Evaluación

- Análisis de todos aquellos factores que podían estar incidiendo durante la implementación del programa.

Objetivos de la Tercera Sesión de Evaluación

- Realizar una evaluación de carácter sumativo sobre los resultados alcanzados en el programa con relación a la consecución de objetivos, contenidos, actividades, etc...

Grado de consecución de los objetivos del programa para el nivel de primero

Las conclusiones fundamentales que podemos establecer en relación a los logros alcanzados en el programa para el nivel de primero, han sido las siguientes:

a) Los objetivos de aprendizaje de las áreas curriculares del programa referidas a Resolución de Problemas, Composición de Números Naturales, Números Fraccionarios, Las Horas, Operaciones y Cálculo, han sido superados por el 75% de los alumnos del nivel de primero excepto en el área curricular «Operaciones y Cálculo».

La figura nº 2 recoge los logros alcanzados en cada una de las áreas curriculares que se han desarrollado durante la implementación del programa.

En cuanto al área curricular del programa referida a **Resolución de Problemas** los objetivos de esa área han sido superados por los alumnos del nivel de primero. En la resolución de problemas de la primera prueba (PR1) los alumnos que mejores resultados han obtenido han sido los del grupo tres y los resultados menos satisfactorios en el grupo uno. En la resolución de problemas de la segunda prueba (PR2) los alumnos de los tres grupos han obtenido resultados muy parecidos a los anteriores; en esta segunda prueba los alumnos del grupo uno han superado los resultados con respecto a los de PR1. La resolución de problemas de la tercera prueba (PR3) ha sido superada por los alumnos del grupo uno y dos. Los alumnos del grupo tres no han alcanzado el

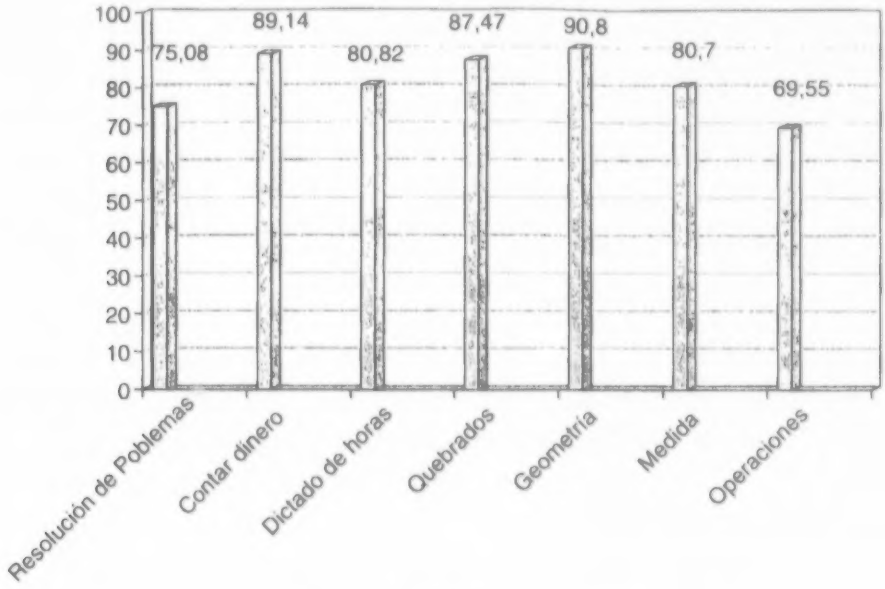


Figura 2
Logros alcanzados en las Áreas Curriculares del Programa en el Nivel de Primero.

nivel mínimo de superación. Los problemas de la cuarta y quinta prueba (PR4) y (PR5) son totalmente escritos sólo se diferencian en el nivel de representación empleado. En la cuarta prueba (PR4) los resultados son muy similares en los alumnos de los tres grupos. Los mejores resultados los han obtenido los alumnos del primer grupo. Los resultados varían sustancialmente en los problemas realizados en la última prueba. En ésta los objetivos del programa referidos a resolución de problemas sólo han sido superados por los alumnos del tercer grupo.

Para el área que va referida a la **Composición de Números Naturales,** los resultados obtenidos muestran que los objetivos del programa relativos a contar dinero con monedas de 1, 5, 10 y 25 pesetas ha sido superados por los tres grupos. Las mayores dificultades han estado en CD2, CD3 y CD4. En estos ítems el uso de las monedas implicaba a la decena.

En relación al conocimiento de las **Horas** los resultados obtenidos muestran que han sido superados los objetivos de aprendizaje referidos a las horas en punto, medias y cuartos. El concepto de medio y cuarto en relación a las horas se trabaja a partir de la introducción que ya se ha realizado previamente en la resolución de problemas.

En cuanto a la **Composición de Números Fraccionarios** se han superado los objetivos del programa que hacen referencia a la utilización de la unidad, el medio y el cuarto. A un nivel manipulativo para la composición de los números fraccionarios sólo se han utilizado círculos, medios círculos y cuartos de círculos.

La introducción que a partir de esta actividad del programa se realiza para la geometría no ha sido trabajada en toda su extensión. En el área de **Geometría** sólo se ha alcanzado el objetivo del programa referido a la discriminación entre cuerpos

redondos y poliedros; actividad que, por otra parte, no ha sido trabajada por todos los profesores. Esta área curricular del programa sólo ha sido evaluada en las dos últimas pruebas realizadas. Los alumnos del grupo uno y tres sólo han realizado la evaluación de esta área en la última prueba.

El apartado del programa destinado al desarrollo de las **Medidas de Peso y Longitud** ha tenido un tratamiento muy puntual. Aunque son conceptos que se desarrollan en la parte media-final del curso su desarrollo en el programa ha sido muy limitado. No se han desarrollado las medidas de peso y de capacidad. Los resultados obtenidos en medidas de longitud han sido satisfactorios; aunque este concepto sólo se haya contemplado en la realización de la última prueba (PR5) si que se han realizado muchos ejercicios en clase con el metro de cartulina que pintaron los niños.

En cuanto al área de **Operaciones y Cálculo** dentro del programa tiene un carácter secundario. Su inclusión en las pruebas obedece fundamentalmente a que los programas de matemáticas recomiendan que los niños ejerciten el cálculo desligado de las situaciones contextuales en las que hay que resolver una situación problemática. Los resultados en esta área han sido del 69,5%.

De estas consideraciones se desprende otras dos conclusiones para el nivel de primero:

- Los objetivos del programa que principalmente han desarrollado los profesores en este nivel educativo pertenecen a las siguientes áreas curriculares: Resolución de Problemas, Composición de Números Naturales y Fraccionarios, Las Horas, Operaciones y Cálculo. Han tenido un menor tratamiento las áreas Medición de Magnitudes y geometría; esta última sólo ha estado referida a la discriminación de figuras geométricas. No se han desarrollado los objetivos del programa correspondientes al valor y tamaño de los ángulos. En el área curricular Sistemas de Numeración, la actividad «numeración casi romana» no se ha trabajado.
- En aquellos objetivos del programa que existen diferencias en los resultados obtenidos por los alumnos las ha establecido el profesor uno (P1) con respecto a los resultados obtenidos por los alumnos del profesor dos (P2) o del profesor tres (P3) para ese mismo objetivo de aprendizaje.

Este resultado se confirma también para la implementación del proceso resolutor: el profesor uno (P1) es el ha marcado diferencias entre los sujetos en cuanto a la comprensión del problema, ejecución y verificación o análisis del resultado obtenido durante el proceso de resolución de problemas.

Grado de consecución de los objetivos del programa para el nivel de segundo

Las diferencias fundamentales establecidas entre el nivel primero y segundo en las distintas áreas curriculares del programa hace referencia a los siguientes aspectos:

1. En cuanto a **Resolución de problemas** se utilizan los mismos conceptos que en el nivel de primero a diferencia de que se realizan mas problemas de tipo mixto y escrito que de tipo gráfico. La numeración que interviene en los problemas es

mas elevada. Los profesores de este nivel no han utilizado la caja de cálculo como material manipulativo para resolver los problemas.

2. En el área **Composición de Números Naturales**, los profesores de segundo, además de las monedas de 1, 5 y 25 pesetas previstas en el programa para la actividad de contar dinero han añadido las de 50 y 100 pesetas.
3. En relación al conocimiento de **Las Horas** los objetivos evaluados coinciden con los especificados en el programa: Horas en punto, medias y cuartos.
4. En la **Composición de Números Fraccionarios** han trabajado la unidad, el medio y el cuarto. Para desarrollar esta actividad han utilizado como materiales básicos el círculo entero, medios círculos y cuartos de círculos. La singularidad de esta actividad en este nivel ha sido que los alumnos han realizado operaciones algebraicas con los números fraccionarios. Han desarrollado la versatilidad de las operaciones con estos números.
5. En este nivel no se han trabajado la discriminación de cuerpos geométricos (cuerpos redondos y poliedros) pero sí la diferenciación en relación al tamaño de los **Ángulos** (grande, mediano y pequeño).
6. No han desarrollado el objetivo de aprendizaje dirigido al conocimiento del valor de los ángulos medidos en minutos horarios.
7. En las **Operaciones y el Cálculo** se han trabajado las operaciones de adición y sustracción con mayor complejidad que en el nivel de primero y se ha introducido a los alumnos en la operación de multiplicar (con una cifra).

Hechas estas diferenciaciones en cuanto a la especificidad que, en el nivel de segundo adquieren los objetivos del programa desarrollado, las conclusiones fundamentales que podemos establecer para este nivel son las siguientes:

- Los objetivos de aprendizaje de las áreas curriculares del programa referidas a Resolución de Problemas, Composición de Números Naturales, Números Fraccionarios, Las Horas, Ángulos, Operaciones y Cálculo, han sido superados por el 75% de los alumnos del nivel de segundo excepto en el área curricular «Resolución de Problemas».

A modo de síntesis el gráfico siguiente ilustra los resultados globales obtenidos para este nivel educativo en las distintas áreas curriculares del programa que se han desarrollado:

- Los objetivos del programa que han desarrollado los profesores en el nivel de segundo pertenecen a las siguientes áreas curriculares: Resolución de Problemas, Composición de Números Naturales y Fraccionarios, Las Horas, Geometría, Operaciones y Cálculo. El área de geometría sólo ha estado referida al valor y tamaño de los ángulos. No se ha desarrollado el área curricular Medición de Magnitudes. La actividad «numeración casi romana» dentro del área «Sistemas de Numeración» no se ha trabajado.
- Para el nivel de segundo, el profesor no ha establecido diferencias entre los alumnos en cuanto a la consecución de los objetivos del programa.

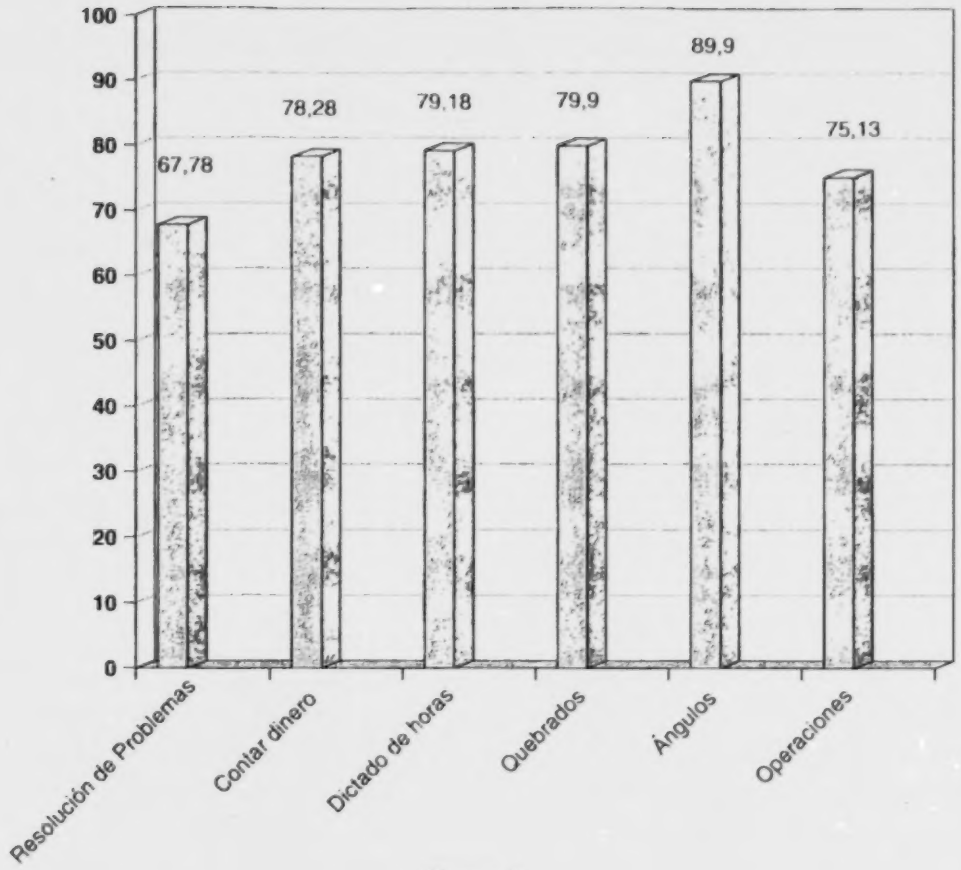


Figura 3
Logros alcanzados en las Áreas Curriculares del Programa en el nivel de segundo.

La visión de los profesores sobre la metodología para la resolución de problemas

Las valoraciones de carácter más interpretativo realizadas por los profesores durante la implementación del programa en tres momentos distintos del mismo, completa desde otras dimensiones la evaluación que en su globalidad hemos realizado del Programa de Iniciación a la Matemática.

Estas valoraciones que realiza el profesorado, se extienden a tres ámbitos diferentes:

- estimación que de forma global han realizado los profesores sobre la implementación del programa: grado de consecución de los objetivos, contenidos, actividades, recursos y aspectos organizativos del programa; factores que han afectado a su funcionamiento, logros principales alcanzados, etc.
- cómo perciben el impacto que el desarrollo del programa ha producido en los alumnos. Valoran que aspectos del programa les ha gustado mas a los alumnos, cuáles menos, si ha habido cambio de actitudes, procesos cognitivos que genera en ellos, etc.

- dimensiones que el programa ha propiciado en los profesores que lo han llevado a la práctica. En este último apartado valoran que aspectos les ha enseñando el programa a ellos, repercusión que ha tenido en su práctica profesional, si el programa ha generado una cambio de actitud sobre lo que es enseñar matemáticas, etc.

Aportaciones del programa a los alumnos según los profesores

El objetivo de esta evaluación es desvelar qué aporta el programa al alumno desde la percepción del profesorado. Este objetivo se concretaría en la siguiente cuestión evaluativa:

¿Qué dimensiones propicia en el alumno el desarrollo del programa?

La información obtenida nos ha posibilitado la categorización de la misma. Se toma como unidad de análisis la unidad conceptual temática significativa. Atendiendo a este criterio las valoraciones que hacen los profesores sobre cómo ha incidido el programa en los alumnos las aglutinamos en las siguientes **dimensiones o categorías**: Cognoscitiva, Actitudinal, Organizativa y Social.

El procedimiento que utilizamos para la elaboración de estas dimensiones fue el siguiente: Una vez recogida la información y seleccionada la unidad de análisis se procedió a la construcción de las categorías o núcleos categoriales que abarcan el conjunto de ideas y temas sobre los que los profesores discuten en las sesiones de evaluación realizadas.

En el **ámbito cognoscitivo**, los profesores consideran que el programa propicia las siguientes dimensiones en los alumnos:

- Razonamiento ordenado de acuerdo a una estructura lógica.
- Desarrolla la atención.
- Posibilita la adquisición de una herramienta válida que puede extrapolar a situaciones problemáticas de la vida real (modelo resolutor).
- El conocimiento experiencial como motor de aprendizaje.
- Capacidad de análisis y síntesis.
- Construcción del discurso matemático que es el que facilita el paso del conocimiento informal a otro formal.
- Lectura de imágenes. Intercambio verbal-visual.

Atendiendo a una dimensión **actitudinal** algunos de los aspectos señalados por los profesores se sitúan en:

- Toma de decisiones
- Análisis de las situaciones
- Autoestima y seguridad en sí mismo
- Investigativas (curiosidad, búsqueda, tenacidad, estimación, planteamiento de hipótesis...)
- Motivación e interés hacía el trabajo
- Participación

En relación a la categoría **aspectos organizativos** que el programa ha generado en los alumnos, a juicio de los profesores, los más importantes han sido:

- Ubicación espacial: todos los niños tienen que mirar a la pizarra cuando se resuelven los problemas.
- La resolución impone un orden en cuanto a quién lo resuelve en cada momento. Se tiene que establecer un sistema rotatorio que permita intervenir cada día que se realizan los problemas a determinados niños.
- Temporalidad: los problemas se hacen únicamente los lunes, miércoles y viernes. El resto de las actividades martes y jueves.

En el **ámbito social**, todos los profesores coinciden en que el efecto principal se traduce en:

- El aula como lugar de encuentro y de construcción del conocimiento. Los problemas se resuelven de forma oral produciéndose una interacción entre niño-profesor. En esta interacción el niño recibe retroalimentación del profesor y de sus iguales (influencias metacognitivas). Una vez resuelto de forma oral, cada niño lo hace individualmente en su cuaderno.

Aportaciones del programa a los profesores que lo han desarrollado

El objetivo de esta evaluación es desvelar qué ha aportado el programa a los profesores que lo han desarrollado. Este objetivo se concretaría en la siguiente cuestión evaluativa:

- **¿Qué dimensiones del programa consideran los profesores que han ejercido una influencia ellos que lo han llevado a la práctica?**

Para el análisis de la información aportada por los profesores sobre los efectos que el programa ha producido en ellos como agentes de su puesta en práctica han utilizado también como unidad de análisis la unidad temática conceptual significativa. De acuerdo a esta unidad de registro el análisis de la información la hemos realizado agrupando las manifestaciones que realizan los profesores en relación al tópico sobre el que se indaga en los mismos núcleos categoriales que señalamos anteriormente.

Algunos de los aspectos señalados por los profesores, muestran las siguientes dimensiones:

- Satisfacción ante la labor docente.
- Una nueva forma de abordar las matemáticas.
- Orden y sistematización.
- Formación en una metodología nueva.
- Conocimiento de una estructura didáctica cíclica.
- Una metodología para la resolución de problemas.
- Conocimiento de una estructura didáctica gradual.
- Ha resultado motivante llevarlo a la práctica.
- Promueve el situarte en una actitud abierta y aperturista.

Dimensiones que también la hemos categorizado con la misma denominación que para las valoraciones que hacen los profesores sobre los efectos que ha producido según ellos en los alumnos: **cognoscitivas, organizativas, actitudinales y sociales.**

Las conclusiones fundamentales que podemos establecer referidos a estos tres ámbitos (el programa en sí, el programa para los alumnos, y el programa en los profesores) son las siguientes:

- El área curricular del programa que más le ha gustado a los profesores y a los alumnos ha sido Resolución de Problemas. La caracterización del proceso resolutor como sistemático, basado en la lógica, conectado con el conocimiento concreto y experiencial del alumno, son algunas de las manifestaciones mostradas por ellos.
- Los logros fundamentales alcanzados a juicio de los profesores son los siguientes: a) construcción de un lenguaje matemático, b) integración de un esquema mental para resolver problemas que hacen extensibles a otras situaciones de la vida cotidiana, c) rigurosidad y sistematicidad que establece el Esquema de Interacción en el proceso resolutor, y d) el programa les enseña a pensar.
- Las actitudes básicas que el desarrollo del programa propicia en el alumno según las valoraciones realizadas por los profesores son: Entusiasmo, de búsqueda, motivación hacía la tarea y autoestima.
- A nivel cognitivo los profesores añaden como aportaciones fundamentales para los alumnos las siguientes: Atención, adquisición de una herramienta conceptual para aplicarla a situaciones problemáticas de la vida real, aprender a pensar en términos de solución de problemas, análisis de situaciones y abstracción.
- Para el profesorado que lo ha llevado a la práctica el programa le ha aportado principalmente una nueva forma de abordar la enseñanza de las matemáticas, mayor rigurosidad y sistematicidad, satisfacción ante la labor docente y una estructura didáctica que ha hecho extensible a otras áreas del currículum escolar teniendo como herramienta el Esquema de Interacción que guía el proceso resolutor.

VALORACIÓN GLOBAL DEL PROGRAMA

A modo de síntesis, expondremos algunos puntos que corresponden a una valoración global del programa así como ciertas recomendaciones para futuras implementaciones del mismo. Finalmente, señalaremos ciertos puntos que requieren aún mayor investigación.

- El programa de Iniciación a la Matemática para niños de primer curso de Educación Primaria ha mostrado resultados satisfactorios en los sujetos de la muestra experimental. Estos resultados se aprecian tanto en el desarrollo del proceso resolutor, en la consecución de los objetivos del programa como en los indicadores cualitativos estimados por los profesores.
- Los resultados obtenidos podemos entenderlo de carácter exploratorio y no como una estimación definitiva; La falta de control sobre la interacción entre selección y tratamiento hace necesaria nuevas aplicaciones del programa.
- Aunque pueda discutirse que los resultados globales obtenidos tengan una traducción en un incremento real de rendimiento matemático es innegable su aportación como herramienta conceptual. El programa de Iniciación a las Mate-

máticas debe considerarse mas un instrumento que facilita la construcción del pensamiento matemático que de incremento de conocimiento matemático.

- El currículum del programa debe ser mantenido en su forma actual, salvo en los puntos que a criterio de los profesores necesitan una revisión.
- Los profesores han detectado sobre los alumnos claros efectos de carácter actitudinal, cognoscitivo, organizativo y social. Para los profesores estos efectos son mas potentes que los meramente cognoscitivos. Estos efectos se han generalizado a otras áreas del currículum escolar.
- El desarrollo del programa ha favorecido un cambio de actitud docente en los profesores que lo han desarrollado. Consideran la experiencia como positiva y los cambios detectados los atribuyen inequívocamente a la intervención del programa.
- La modalidad formativa dispuesta (inicial, de seguimiento y final) ha resultado clave para la implementación del programa.
- El principal problema que presenta su implementación es la resistencia mostrada por los profesores para desarrollarlo acorde al diseño previsto. La ruptura que supone para los profesores con la forma en que tradicionalmente se enseñan las matemáticas.
- El currículum del programa debe ser mantenido en su forma actual, salvo en los puntos que a criterio de los profesores necesitan una revisión.
- Los profesores han detectado sobre los alumnos claros efectos de carácter actitudinal, cognoscitivo, organizativo y social. Para los profesores estos efectos son mas potentes que los meramente cognoscitivos. Estos efectos se han generalizado a otras áreas del currículum escolar.
- El desarrollo del programa ha favorecido un cambio de actitud docente en los profesores que lo han desarrollado. Consideran la experiencia como positiva y los cambios detectados los atribuyen inequívocamente a la intervención del programa.
- La modalidad formativa dispuesta (inicial, de seguimiento y final) ha resultado clave para la implementación del programa.
- El principal problema que presenta su implementación es la resistencia mostrada por los profesores para desarrollarlo acorde al diseño previsto. La ruptura que supone para los profesores con la forma en que tradicionalmente se enseñan las matemáticas, puede esperarse que se den en aplicaciones futuras teniendo en cuenta la duración y los requisitos exigidos.
- El ajuste entre lo teóricamente esperado en el diseño del programa y lo realizado durante su implementación se manifiesta tanto en los resultados mostrados en relación al desarrollo del proceso resolutor como a los logros globales alcanzados con el programa.

RECOMENDACIONES

Sobre el diseño del programa

- Los profesores han destacado que algunas actividades del programa deberían eliminarse y otras actualizarse. Entre las primeras se encuentra la numeración casi romana. Esta actividad los profesores la han considerado de carácter anacrónico. Entre las que habría que actualizar se encuentra la de contar dinero. Ellos proponen la revisión de la actividad en cuanto al uso de las monedas que allí se trabajan.
- En relación al enunciado de los problemas (ya sean gráficos, mixtos o escritos) estiman que hay que revisar el valor adquisitivo de los objetos que intervienen en los problemas planteados. También consideran que no hay una coordinación muy estrecha entre el uso que en los problemas se hace sobre medidas de capacidad y peso y las actividades que trabajan esos conceptos; existen actividades relativas a pesar pero ninguna a medidas de capacidad.
- Estiman que algunos contenidos que se imparten en el programa (números fraccionarios, geometría, ángulos principalmente) exceden los objetivos básicos para este nivel educativo. Esta premisa ha influido para que haya áreas del programa que se han trabajado superficialmente o que no se han tratado.
- Establecer pasos más graduales entre la resolución de problemas de una operación y de dos operaciones; estiman que los cambios entre estos dos tipos de problemas son muy bruscos.
- En la actividad sobre el tamaño de los ángulos consideran que se debería utilizar la denominación usual en los libros de textos: ángulos agudo, recto y obtuso en vez de la propuesta en el programa de grande (obtuso), mediano (recto) y pequeño (agudo).
- Los profesores estiman la conveniencia de extender el diseño del programa hacia el segundo curso de este primer ciclo de Educación Primaria con recursos y actividades adaptados al nivel educativo y en la misma línea pedagógica.
- Rediseñar el tipo y niveles de objetivos del programa para que el modelo teórico de base resulte más accesible y facilite su dominio a los profesores.

Sobre la implementación del programa

- Antes de iniciar una nueva implantación del programa es conveniente estudiar las circunstancias concretas del contexto educativo en el que se va a desarrollar para atenuar o evitar el máximo de problemas institucionales. Los profesores estiman que el factor principal que ha afectado a la implementación efectuada ha sido la falta de tiempo. El comienzo del programa se realizó un mes después de haberse iniciado el curso escolar y la temporalización prevista en el diseño del programa exige la duración completa del curso escolar.
- Profundizar en la modalidad formativa desarrollada para llevar a cabo el programa. La formación del profesorado que lo lleve a sus aulas resulta un factor

clave para su posterior Éxito. Los profesores reclaman la necesidad institucional de realizar grupos de trabajo que potencien la formación en esta metodología.

Sobre la evaluación de la aplicación:

- En futuras aplicaciones del programa sería conveniente diseñar mejores y más ajustados instrumentos de medición para profundizar con mayor precisión en algunos de los logros que los profesores han destacado como fundamentales: el programa como herramienta conceptual que les enseña a pensar; y la validez conceptual y metodológica del Esquema de Interacción como instrumento didáctico que facilita el cambio conceptual en la enseñanza de las matemáticas.
- La evaluación del programa desde una perspectiva longitudinal. Si en el futuro el diseño del programa se hace extensivo al primer ciclo de Educación Primaria, esto nos permitiría valorarlo teniendo como referentes objetivos terminales de ciclo.
- Sería también importante atender en el futuro a hacer más extensiva la aplicación del programa. El desarrollo del programa abierto a más alumnos y profesores con diversidad de características, permite profundizar en algunos aspectos que en esta implementación no han sido tratados.
- Estas recomendaciones generales realizadas para futuras implementaciones del programa abren además **líneas de investigación** para profundizar en la intervención iniciada con el Programa de Iniciación a la Matemática. Estas líneas de trabajo nos parecen muy útiles para mejorar los resultados del programa o incluso para abrir nuevas perspectivas.

BIBLIOGRAFÍA

BAROODY, A. (1994): El pensamiento matemático de los niños. Un marco evolutivo para maestros de preescolar, ciclo inicial y educación especial. Aprendizaje-Visor: Madrid.

CASTRO, E.; RICO, L. y GIL, F. (1992): «Enfoques de investigación en problemas verbales aritméticos aditivos». *Enseñanza de las Ciencias*, 10 (3), 243-253.

CASTRO, E. (1995): *Niveles de comprensión en problemas verbales de comparación multiplicativa.* Granada: Comares.

CASTRO, E.; RICO, L.; GUTIÉRREZ, J.; CASTRO, E.; SEGOVIA, I.; MORCILLO, N.; FERNÁNDEZ, F.; GONZÁLEZ, E. y TORTOSA, A. (1996): «Evaluación de la resolución de problemas aritméticos en Primaria». *Revista de Investigación Educativa*, Vol. 14, nº 2.

COLAS, P. y REBOLLO, M.A. (1993) (2ª Ed.). *Evaluación de programas. Una guía práctica.* Sevilla: Kronos.

GONZÁLEZ RAMÍREZ, T. (1997): «Evaluación de un Programa de Iniciación a las Matemáticas a través de la Resolución de Problemas». Tesis Doctoral Inédita. Universidad de Sevilla.

GONZÁLEZ URBANEJA, P.M. (1991): «Historia de la matemática: Integración cultural de las matemáticas: Génesis de los conceptos y orientación de su enseñanza». Enseñanza de las Ciencias, 9(3), pp. 281-289.

KLINE, M. (1978). *El fracaso de la matemática moderna*. Siglo XXI: Madrid.

LAPOINTE, A. E., MEAD, N.A. y PHILIPS, G.V. (1989). *A world of differences*. Princenton, NL, Educational Testing Service (Trad. cast: *Un mundo de diferencias*. Madrid, CIDE).

LOPES, B. y COSTA, N. (1996): «Modelo de enseñanza-aprendizaje centrado en la resolución de problemas: Fundamentación, presentación e implicaciones educativas». Enseñanza de las Ciencias, 14 (1), pp. 45-61.

MASINGILA, J. (1996). Learning from students out-of-school mathematics practice. *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education*. Valencia, 8-12 de Julio.

MAZA, C. (1995). *Aritmética y representación. De la comprensión del texto al uso de materiales*. Barcelona: Paidós.

MINISTERIO DE EDUCACIÓN Y CIENCIA (1989): Diseño Curricular Base. Educación Primaria. Madrid.

PUTNAM, R.T.; LAMPERT, M. y PETERSON, P.L. (1990). Alternative perspectives on knowing mathematics in elementary schools. En C.B. Cazden, *Review of Research in Education*, 16 Washington. AERA, 57-150.

ROGOFF, B. AND LAVE, J. (eds.) (1995). *Everyday cognition: its development in social context*. Cambridge, MA, Harvard University Press.

SAXE, G.B.; GUBERMAN, S.R.; GEARHART, M. (1993). Social processes in early number development. *Monographs in the Society for Research in Child Development*, 52 (serial n. 216).

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 201-220

ANÁLISIS Y VALIDACIÓN DE UNA ESCALA PARA MEDIR LA CONDUCTA EXPLORATORIA

Trinidad Donoso, Pilar Figuera, Mercedes Torrado
Universidad de Barcelona

RESUMEN

La importancia de la conducta exploratoria en la gestión del proyecto profesional-vital está ampliamente documentada en la literatura sobre el desarrollo de la carrera en las dos últimas décadas. Si bien la conducta de exploración ocurre a lo largo de todo el desarrollo, uno de los momentos donde adquiere mayor relevancia se produce durante la etapa de adolescencia y juventud. La importancia de esta variable contrasta con la ausencia de instrumentos para evaluar la conducta exploratoria de adolescentes y jóvenes en procesos de transición al mercado laboral, en un contexto de empleo que dificulta la entrada de los más jóvenes así como la consolidación de su proyecto profesional. Esta necesidad nos llevó a adaptar el Cuestionario «Career Exploration Survey» (Blustein; 1988) al contexto español. El artículo describe el proceso seguido y los análisis estadísticos de la escala que documentan la adaptación de la misma.

ABSTRACT

The importance of the career exploration along with the management of the vital-professional project has been broadly documented in the careers' literature during the last two decades. Although the career exploration occurs during the entire personal development, one of the most outstanding phases is during the adolescence and the youth stages. The importance of these phases contrasts with the lack of tools to be able to evaluate the career exploration among adolescents and the young during their transitional steps into the labour market. A market

where the scarcity of job employment offers holds in check the careers of the youngest as well as the consolidation of their professional success. The need for the proper tools lead us to readapt the «Career Exploration Survey» (Blustein; 1988) to the Spanish context. The article describes the followed procedures as well as the statistical analysis of the scales which support the conversion.

1. INTRODUCCIÓN

La investigación sobre el desarrollo de la carrera pone de manifiesto el efecto de la conducta exploratoria sobre los procesos de elección e inserción socio-profesional de jóvenes y adultos. El corpus de investigación creciente sobre los procesos de planificación de los itinerarios educativos y profesionales aporta datos relevantes sobre el valor fundamental de la conducta desarrollada por la persona, tanto haciendo referencia al tipo de conducta como a las variables relacionadas, que ayudan a comprender el propio constructo.

En los procesos de toma de decisiones, la conducta exploratoria aparece como precursora de la cristalización, elección e implementación del proyecto profesional. Así, investigaciones realizadas sobre la influencia de las actividades de exploración de los jóvenes (prácticas laborales, investigación del mercado laboral, autoexploración, trabajo voluntario), resaltan su efecto directo sobre las tareas de desarrollo de la carrera, al contribuir a la «autorreflexión» y el «proceso de decisión», permitiendo la «elaboración de planes o proyectos profesionales» más realistas (Blustein, 1989; Altman y Sedlacek, 1991; Hoyt, 1995) y la «cristalización de una identidad vocacional» (Heppner y Cook, 1991). Estos resultados incrementan los recursos personales para gestionar los procesos de transición, ya que su efecto se produce sobre la efectividad de los procesos decisorios y la confianza en la decisión, considerándose un componente esencial de la madurez de carrera (Super, 1994).

Durante la fase de transición al mercado laboral, la necesidad de dar dirección al proyecto profesional-vital, a través de tareas de carrera como «elegir un campo ocupacional», «identificar vías de inserción» o «generar redes de contacto», justifican la importancia de la conducta exploratoria. La investigación ha puesto de manifiesto que aquellas personas más implicadas en la exploración de sí mismas y del ambiente tienen una mayor conciencia de sus capacidades, valores y aspiraciones, realizan opciones más congruentes con sus posibilidades y preferencias, tienen un mayor conocimiento de las vías de inserción e incrementan sus posibilidades de transición; de tal manera que la inserción es directamente proporcional a la conducta desarrollada antes y durante el proceso de transición al mercado laboral (véase revisión de investigaciones en Figuera, 1996). Diversos estudios, realizados en el contexto internacional y en nuestro país, encuentran evidencia de su asociación con los procesos de entrada y ajuste ocupacional (Stefy, Shaw y Noe, 1989; Blanch, 1990), siendo también un precursor de la satisfacción laboral (Stumpf, Austin y Hartman, 1984; Taylor, 1985).

Si desde una perspectiva global puede afirmarse que una mayor actividad de búsqueda incrementa las posibilidades de inserción, el nivel de efectividad en la

predicción del logro de empleo depende de variables como las características de la actividad o tipo de afrontamiento realizado, medido a través de indicadores como sistematicidad, proacción, u orientación a la meta (Risk, 1987), así como el grado de autonomía personal y el grado de flexibilidad en la definición del plan profesional (Benoit-Guilbot, 1990).

La investigación actual sobre los procesos de transición ha obtenido evidencias de que diversas personas con cirscunstancias y oportunidades similares pueden conseguir variados niveles de éxito debido a factores personales. Buena parte de ellos quedan resumidos en el constructo de exploración de la carrera profesional. La conducta exploratoria constituye pues un constructo esencial del desarrollo de la carrera y uno de los focos prioritarios de atención en la intervención para el desarrollo de la gestión del proyecto profesional (Pereira, 1995; Hoyt, 1995; Rodríguez y al., 1995).

2. EL INSTRUMENTO ORIGINAL

2.1. Antecedentes en la medición de la conducta exploratoria

Los antecedentes en el estudio de la exploración de la carrera profesional se inician desde la dimensión conductual y en la medición de la frecuencia y variedad de los actos exploratorios (información, búsqueda de empleo, etc.). Los primeros estudios dentro del ámbito de la psicología experimental entendían la conducta en términos de refuerzo (Berlyne, 1960), cuya función era obtener información adicional del medio. Desde la psicología evolutiva el concepto de exploración se utilizará para etiquetar una de las etapas de la vida donde la búsqueda de una definición profesional es más intensa. Serán los teoricos del desarrollo vocacional, a partir fundamentalmente de los trabajos de Jordaan y Super, quienes definan la exploración como una tarea implícita al desarrollo de la carrera y se investigue como base de la madurez vocacional en adolescentes, primero, y ampliada con posterioridad a todo el ciclo vital. En este contexto evolutivo, Jordaan (1963) aportará una delimitación estructural del constructo, identificando los elementos básicos del comportamiento de exploración de la carrera conceptualizado como conducta mental o física que permite obtener información acerca de uno mismo o de su medio con el fin de tomar decisiones, concluir, solucionar, construir hipótesis o para cambiar, sentirse estimulado o responder a los retos. Ligado todavía a una concepción conductual, el análisis de los determinantes permite entreveer un acercamiento a los actuales modelos explicativos de la conducta. Jordaan introduce en su modelo toda una serie de factores facilitadores o inhibidores de la conducta de exploración tanto externos (estimulación del ambiente, tensiones sociales,...) como internos (capacidad de tolerar la tensión, la incertidumbre o la ambigüedad y a la frustración; la objetividad y aptitud para formar juicios; la imagen y confianza en uno mismo; la automonía, la madurez y la apertura al medio).

Los enfoques cognitivos traducirán estos elementos dentro de un modelo más global de exploración donde los componentes conductuales son efecto y, a la vez causa, de los procesos cognitivos y de las reacciones afectivas a los mismos. Este conjunto de variables determina la dirección y las consecuencias de la acción en los

logros personales, como son los procesos de gestión del proyecto profesional (Stumpf, Colarelli y Hartman, 1983; Stumpf, Austin y Hartman, 1984; Stumpf y Lockhart, 1987; Stumpf, Brief y Hartman 1987).

La línea actual del desarrollo de la carrera conceptualiza la conducta exploratoria como aquellas accciones cognitivas y conceptuales que la persona realiza con la finalidad de incrementar el conocimiento de sí misma y del contexto social, educativo y laboral que le permiten optimizar sus recursos personales para una mejor gestión de su proyecto profesional-vital. La investigación reciente integra el estudio del proceso de exploración de la carrera en el marco sociopersonal donde se desarrolla (Taveira y otros, 1998). Las investigaciones han identificado la interacción de los elementos motivacionales e individuales de la exploración, y tratan de situar los determinates personales (características individuales) y sociales (estructura de oportunidades) en un modelo más comprensivo (en esta línea destacan los trabajos de Lent y col., 1994, Lent y Hackett, 1994; Lent y Brown, 1996 y b; Lent, 1999)

2.2. Modelo teórico de la escala

La escala se basa en las aportaciones de Stumpf, Colarelli y Hartman quienes presentan un modelo integral de la exploración profesional que incluye tanto elementos conductuales como cognitivos acerca de la misma conducta (Stumpf, Astin y Hartman, 1984; Stumpf, Brief y Hartman, 1987; Stumpf y Lockhart, 1987; Blustein, 1988, 89, 92). En base, por un lado a los trabajos pioneros de Berlyne (1960) y Jordaan (1963) y, por el otro, a los postulados básicos de las concepciones cognitivo-conductuales (Bandura; Folkman y Lazarus), este modelo representa una de las aportaciones teóricas más representativas del estudio del constructo.

El modelo tiene como elemento central el proceso de exploración, en su dimensión conductual, que interactúa con dos núcleos de variables: las creencias (cogniciones acerca de uno mismo y de su entorno) y las reacciones (de satisfacción y de estrés). Entroncando con los enfoques cognitivos de la conducta, como el modelo de afrontamiento de Lazarus y Folkman y la teoría socio-cognitiva de Bandura, el modelo admite que el comportamiento es un precursor importante de los logros, pero entiende que la actividad exploratoria no puede limitarse al estímulo o demanda externa, sino que se genera en el marco de la interpretación personal de las demandas ambientales.

Una serie de principios o elementos son inherentes a este enfoque:

1. A través de los itinerarios o carreras profesionales, las personas se encuentran con eventos que tienen que afrontar. El concepto de exploración es, pues, un proceso de afrontamiento de uno de estos eventos. La exploración profesional es el conjunto de acciones cognitivas y conductuales que de una manera intencional permiten el acceso a la información sobre trabajos, ocupaciones u organizaciones.
2. La conducta (dónde, cómo, cuánto y qué) es el resultado de la interacción de procesos cognitivos, conductuales y afectivos. Como elemento clave del proceso de exploración, la conducta no sólo interactúa con los procesos cognitivos (creen-

CUADRO 1
MODELO ESTRUCTURAL DE LA EXPLORACIÓN PROFESIONAL. DIMENSIONES CONCEPTUALES

LA DIMENSIÓN CONDUCTUAL →	**EL PROCESO DE EXPLORACIÓN**
Las conductas que una persona lleva a cabo durante el proceso de exploración profesional	Las subdimensiones del proceso se concretan en: • Dónde se realiza la exploración • Cómo se explora • Cuánto se explora • Qué se busca
LA DIMENSIÓN COGNITIVA →	**EXPECTATIVAS Y CREENCIAS**
Las conductas y percepciones que tiene la persona sobre la exploración profesional	Las subdimensiones del proceso se concretan en: • percepción sobre las condiciones del mercado de trabajo • creencias en los métodos de exploración • importancia que se le dan a las metas profesionales
LA DIMENSIÓN AFECTIVA →	**REACCIONES A LA EXPLORACIÓN**
Los sentimientos que se experimentan durante el proceso de exploración profesional	Las subdimensiones del proceso se concretan en: • grado de satisfacción con las conductas realizadas • ansiedad vivida en el proceso.

cias) y afectivos (reacciones), sino que, además, tiene un peso importante en los logros de la carrera y constituye un elemento central del proceso de inserción (como bien señalan Stumpf, Austin y Hartman, 1984).

3. El proceso de exploración incluye tres dimensiones básicas: la dimensión conductual, la dimensión afectiva o reacciones a la exploración y la dimensión cognitiva del proceso o conjunto de creencias de la persona. Entre ellas se producen influencias causales bidireccionales, tal como aparece en el modelo (véase Donoso Figuera y Rodríguez, 1994).

3. LA ESCALA DE CONDUCTA EXPLORATORIA (ESCE)

El cuestionario CES (Career Exploration Survey) en su versión original se ajusta a la estructura del modelo de Stumpf, Colarelli y Hartman (1981), es un cuestionario multidimensional compuesto por 59 ítems que pretende evaluar la conducta explora-

CUADRO 2
SUBDIMENSIONES DE LA ESCALA Y EJEMPLOS DE ÍTEMS CORRESPONDIENTES

Proceso de exploración	Ejemplo de preguntas
EXPLORACIÓN AMBIENTAL: Extensión de la exploración de la profesión examinando ocupaciones, trabajos y organizaciones en los últimos tres meses. (ítems 14-19)	«cuál de las siguientes actividades has realizado durante los tres últimos meses: — Obtener información sobre empleos
EXPLORACIÓN INTENCIONAL Y SISTEMÁTICA: Extensión con que uno adquiere información sobre uno mismo y el ambiente de una manera sistemática. (ítems 11-13)	(...) — Adquirir experiencias en diferentes actividades laborales.
AUTO-EXPLORACIÓN: Extensión de la exploración de la carrera que comprende auto-valoración y retrospección en los últimos tres meses. (ítems 20-24)	(...) — Reflexionar sobre mi rendimiento académico
NÚMERO DE OCUPACIONES CONSIDERADAS: Número de diferentes áreas ocupacionales en las que uno está adquiriendo información. (ítem 25)	— busco información en un ámbito ocupacional versus en muchos ámbitos ocupacionales.
FRECUENCIA: Promedio de veces por semana que uno busca información profesional en un periodo de dos meses. (ítem 29)	— Cuántas veces por semana has intentado obtener información...
CANTIDAD DE INFORMACIÓN: Cantidad de información adquirida sobre ocupaciones, trabajos, organizaciones y uno mismo. (ítems 1-3)	— ¿Qué grado de información tienes sobre lo que hacen las personas que trabajan en tu profesión?
FOCO: Seguridad de la preferencia por una ocupación particular, trabajo y organización. (ítems 8-10)	— Hasta qué punto estás seguro de conocer el tipo de trabajo que más te conviene

Reacciones a la exploración	Ejemplo de preguntas
SATISFACCIÓN CON LA INFORMACIÓN: Satisfacción con la la información obtenida respecto a ocupaciones, trabajos y organizaciones relativas a los propios intereses, capacidades y necesidades (ítems 4-7)	«Hasta qué punto estás satisfecho con información de que dispones sobre»: — el empleo concreto sobre el que estás interesado
TENSIÓN EXPLORATORIA: Cantidad de tensión no deseable en relación a otros eventos significativos de la vida. (ítems 44-46)	«Qué ansiedad te ha ocasionado/aria»: — enfrentarme a procesos de selección.
TENSIÓN DECISORIA: Cantidad de tensión no deseable, en relación a otros eventos significativos de la vida. (ítems 47-50)	(...) — decidir lo que quiero hacer.

Expectativas	Ejemplo de preguntas
PERSPECTIVA PROFESIONAL: Cuán favorables son las posibilidades que uno busca en un área profesional. (ítems 26-28)	«Indica cuáles son tus expectativas de conseguir un empleo»: — en tu campo profesional
INSTRUMENTALIDAD DE LA BÚSQUEDA EXTERNA: La probabilidad que explorar el ambiente sobre oportunidades profesionales conducirá a obtener metas profesionales. (ítems 37-39)	«Importancia de las siguientes actividades de cara a conseguir tus metas profesionales»: — Conversar con amigos y conocidos acerca del mundo laboral.
INSTRUMENTALIDAD DE LA BÚSQUEDA INTERNA: Probabilidad de que reflexionar retrospectivamente sobre conductas profesionales pasadas conducirá a conseguir las metas profesionales. (ítems 34-36)	(...) — Aprender más acerca de mí mismo.
INSTRUMENTALIDAD DEL MÉTODO: Probabilidad de que la exploración profesional intencional y sistemática conducirá a conseguir las metas profesionales. (ítems 30-33)	(...) — Planificar cuidadosamente la búsqueda de empleo.
IMPORTANCIA DE OBTENER UN PUESTO DE TRABAJO PREFERIDO: Grado de importancia que se da al hecho de obtener una profesión preferente. (ítems 40-43)	«Qué importancia tiene para ti en este momento»: — Ocupar un determinado nivel profesional.

toria de acuerdo a las tres dimensiones del modelo presentado (creencias, proceso y reacciones) que a su vez agrupan otras subdimensiones.

La versión utilizada para realizar la adaptación es de Blustein (1988) y la traducción y adaptación al contexto español fue realizado por Donoso, Figuera y Rodríguez (1994) reduciendo el número de ítems original de 59 a 50. Los criterios para realizar las reducciones y los cambios en los ítems y dimensiones correspondientes puede consultarse en la publicación de las autoras mencionadas anteriormente y la publicación de la escala en Rodríguez Moreno, (1999).

En las tablas siguientes se especifican las subdimensiones del modelo junto con un ejemplo de preguntas y entre paréntesis los ítems que corresponden a cada dimensión y su organización en la escala adaptada.

4. EL PROCESO DE ADAPTACIÓN DE LA ESCALA AL CONTEXTO ESPAÑOL

4.1. Descripción y análisis de los resultados

Para poder validar la adaptación realizada de la escala ESCE se procedió durante el curso académico 1997-98 a efectuar un pase piloto de la escala a un total de 125 sujetos de Ciclos Formativos (CF) de Grado Medio y de Grado Superior y de Formación

Profesional se segundo grado (FP II) de centros públicos de la ciudad de Barcelona y comarca. Para ello dispusimos de la colaboración de los tutores y tutoras responsables de los respectivos grupos.

La descripción de los resultados de la escala en esta muestra se presentan en el artículo *La conducta exploratoria: concepto y aplicaciones en orientación profesional* a cargo de Rodríguez, M.L, Buisan, C y Sandin, M.P. (en prensa).

En el presente artículo únicamente abordaremos la descripción y análisis interno de los ítems, para de este modo validar el constructo teórico propuesto y validar internamente la consistencia de los ítems. Para ello se se ha utilizado el programa estadístico SPSS-Win (versión 8.0) en el cálculo de todas las pruebas estadísticas: Análisis factorial y análisis de la consistencia interna.

4.2. Análisis factorial

Para poder confirmar las dimensiones y subdimensiones de la escala ESCE, se ha utilizado el análisis factorial como técnica analítica multivariada que nos permitirá reducir el conjunto de los 50 ítems propuestos en nuestra adaptación, a unos pocos factores o dimensiones.

Dentro de las posibilidades que permite este análisis se ha optado por aplicar el análisis de los componentes principales (ACP), por ser el más apropiado para el estudio de escalas cuya estructura y dimensionalidad ya habían sido objeto de análisis previos en versiones anteriores, como es el caso de la presente escala. Este análisis consiste en condensar la matriz de correlaciones entre las variables en unos «componentes principales» de la variabilidad total; es decir, el ACP es una técnica estadística que permite transformar un conjunto de variables intercorrelacionadas en otro conjunto de variables no correlacionadas denominadas factores (Bisquerra, 1989).

Previamente al análisis factorial se han confirmado que las condiciones de las variables y valores son apropiados para este tipo de análisis mediante el cálculo del *determinante de correlación, la prueba KMO* (Kaiser-Meyer-Olkin) y *el test de esfericidad de Barlett.*

En la siguiente tabla se presentan los resultados del análisis factorial de la matriz rotada, mediante la aplicación del método VARIMAX. Los ítems no se han ordenado según sus saturaciones en los 15 factores resultantes, sinó según el orden de aparición en la escala, de esta forma las saturaciones de los factores no están ordenadas; con la finalidad de facilitar la lectura de adecuación o no a las 15 subdimensiones contempladas en el instrumento y que componen las tres dimensiones (PROCESO DE EXPLORACIÓN, REACCIONES A LA EXPLORACIÓN Y EXPECTATIVAS) Para ello se separan cada una de ellas con una línea gruesa.

TABLA 1
ANÁLISIS FACTORIAL DE LA ESCALA ESCE

	FACTORES														
ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Item1				0,5570											
Item2				0,4889					0,3924						
Item3													0,7608		
Item4									0,7107						
Item5									0,8203						
Item6											0,6189				
Item7									0,4643						
Item8				0,7116											
Item9				0,7254											
Item10				0,7708											
Item11										0,8631					
Item12	0,3549			0,3269					0,3550	0,3965	-0,3249				
Item13										0,5478					
Item14										0,6471					
Item15	0,8212														
Item16	0,7604														
Item17	0,5486														
Item18	0,6852														
Item19	0,6542														
Item20		0,6898													
Item21		0,7509													
Item22		0,7512													
Item23		0,7483													
Item24		0,5893													
Item25			0,5297												
Item26							0,8044								
Item27							0,8185								
Item28							0,5847								
Item29	0,4662													0,4035	
Item30						0,7154									
Item31						0,7300									
Item32						0,7519									
Item33			0,4777			0,3753									
Item34			0,7358												
Item35			0,5559												
Item36			0,7876												
Item37															0,6083
Item38												0,6614			
Item39												0,6961			
Item40								0,6270							
Item41								0,8235							
Item42								0,6461							
Item43								0,5692							
Item44					0,5886										
Item45			0,4939						-0,3514						
Item46														0,6848	
Item47					0,8655										
Item48					0,8279										
Item49					0,4755							0,4659			
Item50											0,7031				
%Var	18,3%	12,3%	6,1%	5,2%	4,4%	3,9%	3,5%	3,3%	2,9%	2,7%	2,5%	2,3%	2,3%	2,2%	2,0%

BEST COPY AVAILABLE

Este primer análisis factorial confirma la existencia de 15 dimensiones que coinciden, en la gran mayoría, a las subdimensiones contempladas en el constructo teórico. Con estos 15 factores se explica el 74% de la varianza. Existen, sin embargo, algunas excepciones. En concreto el ítem 3 (*Dificultades para saber mi futuro profesional*) constituye él solo el factor 13, desligándose de la subdimensión **Cantidad de información**; el ítem 6 (*Los tipos de organizaciones laborales y empresas que más pueden adecuarse a tus necesidades personales*) es junto con el ítem 50 (*Decidirme sobre qué organizaciones o empresas pueden convenirme más*) los dos elementos que engloba el factor 11, lo mismo sucede con los ítems 13 y 14. El resto de factores, tal y como se puede observar en el cuadro resumen coinciden con las subdimensiones.

Por otro lado ítems como el 12, el 29, el 45 y el 49 no se destacan por tener una saturación en algún factor concreto, si no que más bien se distribuyen entre varios con niveles de saturación menores a 0,466. La confección de cuadro 3 resume los resultados comparándolos con el constructo que se pretende confirmar.

CUADRO 3
RESUMEN COMPARATIVO DE LAS DIMENSIONES TEÓRICAS Y LAS RESULTANTES EN EL ANÁLISIS FACTORIAL DE TODA LA ESCALA

	CONSTRUCTO TEÓRICO		DIMENSIONES CONFIRMADAS	
	Subdimensiones	Ítems contemplados	Ítems confirmados	FACTORES
PROCESO DE EXPLORACIÓN	Exploración Ambiental	Del 13 al 19	Del 15 al 19	FACTOR 1
	Autoexploración	Del 20 al 24	Del 20 al 24	FACTOR 2
	N° Ocupaciones consideradas	El ítem 25		
	Intencionalidad Sistemática	Del 11 al 12	El ítem 11	FACTOR 10
	Frecuencia	El ítem 29		
	Cantidad de información	Del 1 al 3	Del 1 al 2, del 8 al 10	FACTOR 4
	Foco	Del 8 al 10		
REACCIONES A LA EXPLORACIÓN	Satisfacción con la información	Del 4 al 7	Del 4 al 5 y el 7	FACTOR 9
	Tensión de exploración	Del 44 al 46	El ítem 46	FACTOR 14
	Tensión de decisión	Del 47 al 50	Del 47 y 48	FACTOR 5
EXPECTATIVAS	Perspectiva de empleo	Del 26 al 28	Del 26 al 28	FACTOR 7
	Instrumentalidad de la búsqueda externa	Del 37 al 39	El 38 y 39	FACTOR 12
	Instrumetalidad de la búsqueda interna	Del 34 al 36	Del 34 al 36	FACTOR 3
	Instrumentalidad del método	Del 30 al 33	Del 30 al 33	FACTOR 6
	Importancia obtener posición deseada	Del 40 al 43	Del 40 al 43	FACTOR 8

Se ha realizado también un análisis factorial agrupando los ítems según las tres dimensiones para confirmar la existencia de factores diferenciados dentro de ellas.

La dimensión PROCESO DE EXPLORACIÓN está explicada en un 60,4% por cinco factores. De los cuales dos (factor 2 y 3) coinciden con la subdimensión del constructo (**autoexploración y foco**).

A la luz de los resultados se observa que el ítem 3 (*Tengo dificultades para saber en que trabajaré yo en el futuro,...*) tiene su mayor peso en la subdimensión **Foco** y no en **Capacidad de información,** no obstante presenta una saturación baja con respecto al resto de ítems de esta dimensión; que el ítem 12 (*Buscar los medios para profundizar en el conocimiento de mis capacidades*) está muy repartido entre los factores 3 y 4 y por último que el ítem 25 (*En estos momentos estoy buscando información principalmente en una ocupación/ Busco información sobre muchos campos ocupacionales*) está repartido entre el factor 1 y 4.

TABLA 2
ANÁLISIS FACTORIAL POR DIMENSIONES. DIMENSIÓN PROCESO DE EXPLORACIÓN

Subdimensión	ITEMS	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Capacidad de información	ITEM 1					0,84128
	ITEM 2					0,76496
	ITEM 3	0,23312		0,37640	0,23600	
Foco	ITEM 8			0,76007		
	ITEM 9			0,67863		
	ITEM 10			0,80511		
Explointención	ITEM 11				0,83230	
	ITEM 12			0,42482	0,44454	
Exploración Ambiental	ITEM 13				0,60866	
	ITEM 14				0,69357	
	ITEM 15	0.83382				
	ITEM 16	0,74658				
	ITEM 17	0,56975				
	ITEM 18	0,62667				
	ITEM 19	0,68493				
Autoexploración	ITEM 20		0,72428			
	ITEM 21		0,74657			
	ITEM 22		0,77034			
	ITEM 23		0,78726			
	ITEM 24		0,63209			
N° de Ocupaciones	ITEM 25	0,33989			0,37099	
Frecuencia	ITEM 29	0,40907				
	% VAR	30,1%	11%	8,9%	5,5%	4,9%

En la dimensión REACCIONES A LA EXPLORACIÓN son tres los factores que aparecen, de los cuales dos coinciden en su totalidad a la subdimensión contemplada. La única discrepancia radica en el ítem 44 (*Buscar información sobre mis preferencias profesionales*) ya que es el único que no se adecúa al factor correspondiente, es decir, su saturación es mayor en el factor 1 donde se engloba la totalidad de preguntas de la subdimensión Tensión en la decisión, aunque su peso está repartido también en el factor 3.

TABLA 3
ANÁLISIS FACTORIAL POR DIMENSIONES. DIMENSIÓN REACCIONES A LA EXPLORACIÓN

Subdimensión	ITEMS	Factor 1	Factor 2	Factor 3
Satisfacción con la información	ITEM 4		0,75160	
	ITEM 5		0,84818	
	ITEM 6		0,69128	
	ITEM 7		0,66073	
Tensión en la exploración	ITEM 44	0,52408		0,47915
	ITEM 45			0,84909
	ITEM 46			0,59025
Tensión en la decisión	ITEM 47	0,80713		
	ITEM 48	0,82039		
	ITEM 49	0,71618		
	ITEM 50	0,63902		
	% VAR	29,1%	20,9%	9,4%

El estudio de la dimensión EXPECTATIVAS presenta la agrupación de sus ítems en cinco factores, coincidiendo casi en su totalidad a las cinco subdimensiones de esta dimensión de la escala. Existen, no obstante, dos ítems, el 33 (*Autoevaluarme con el fin de encontrar un trabajo adecuado para mi*) y el 37 (*Obtener información sobre el mercado laboral y sobre oportunidades de trabajo en mi área profesional)* que se ubican en subdimensiones diferentes.

A modo de conclusión el análisis factorial por dimensiones ha confirmado la consistencia de las subdimensiones existentes en cada una de las dimensiones contempladas en la Escala. Los ítems 3 y 25 son los únicos que presentan una dispersión entre los diferentes factores. Otros ítems como el 12 y el 44 su peso explicativo está repartido casi en igualdad entre dos factores. Y los ítems 13 y 14 de la dimensión Proceso de exploración y los ítems 33 y 37 de la dimensión Expectativas son los únicos que no se encuentran en los factores correspondientes.

De los resultados del análisis factorial de la escala ESCE tanto al nivel de toda la escala como por dimensiones se pueden extraer las siguienes conclusiones:

- Se puede afirmar que en general las dimensiones saturan en factores independientes, de lo que se deduce que el análisis de la conducta exploratoria debe

TABLA 4
ANÁLISIS FACTORIAL POR DIMENSIONES. DIMENSIÓN EXPECTATIVAS

Subdimensión	ITEMS	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Perspectiva Profesional	ITEM 26			0,82782		
	ITEM 27			0,91041		
	ITEM 28			0,76364		
Instrumentalidad	ITEM 30		0,81973			
del método	ITEM 31		0,81889			
	ITEM 32		0,64427			
	ITEM 33	0,61120				
Instrumentalidad de	ITEM 34	0,83024				
la búsqueda interna	ITEM 35	0,64075				
	ITEM 36	0,87325				
Instrumentalidad de	ITEM 37		0,63250			
la búsqueda externa	ITEM 38					0,75262
	ITEM 39					0,72777
Importancia obtener trabajo	ITEM 40				0,62480	
	ITEM 41				0,84928	
	ITEM 42				0,66912	
	ITEM 43				0,61691	
	% VAR	31,2%	13,6%	9,9%	7,1%	6%

tener en cuenta estas dimensiones y subdimensiones como aspectos relevantes en el estudio del constructo.

- Que el ítem 3 manifiesta un comportamiento diferencial con respecto al resto de los ítems tanto a nivel de toda la escala (constituyendo un sólo factor) como a nivel de dimensión Proceso de Exploración (presentado una dispersión de saturación en los cinco factores resultantes).

4.3. Análisis de la validez interna

Para constatar la consistencia interna de la escala ESCE se han calculado varios índices que en su conjunto nos ayudarán a confirmar o no nuestro objetivo: ***que la escala es consistente y por lo tanto fiable.***

Los resultados que se presentan hacen referencia al estudio de las dimensiones y subdimensiones que componen la Conducta exploratoria y a los ítems que constituyen la escala.

Según la tabla 4 todas las intercorrelaciones entre dimensiones son positivas y significativas. Lo cual parece evidenciar que todas las dimensiones miden el mismo

constructo. Ahora bien, si analizamos las correlaciones de cada dimensión observamos que existe una elevada correlación interna con algunos matices pero que en cualquier caso todos los valores obtenidos presentan una significación estadística (1%).

- en la dimensión PROCESO DE EXPLORACIÓN las correlaciones entre sus siete subdimensiones son todas significativas oscilando entre 0,5115 (Autoexploración) y 0,7851 (Exploración ambiental).
- en la dimensión REACCIONES A LA EXPLORACIÓN la subdimensión que presenta una correlación menor pero significativa con un margen de error del 1% es *Satisfacción con la información* (0,4302), asimismo, la que presenta mayor grado de relación es *Tensión de decisión* (0,8403).
- en la última dimensión del constructo, EXPECTATIVAS la mayoría de subdimensiones presentan un grado similar de relación (valores alrededor 0,75), únicamente las dimensiones *Perspectivas profesionales* e *Importancia de obtener un trabajo* presentan valores menores (0,4858 y 0,6580 respectivamente).

TABLA 5
INTERCORRELACIONES ENTRE DIMENSIONES DE CONDUCTA EXPLORATORIA

DIMENSIONES	DIMENSIONES			TOTAL
	Proceso de Exploración	Reacciones a la exploración	Expectativas	
Proceso de Exploración		0.3317**	0,4276**	0,7270**
Reacciones a la exploración			0,5791**	0,8018**
Expectativas				0,8423**

Con respecto al análisis de los ítems, se ha procedido, por un lado, a calcular el grado de relación existente entre los ítems y el total de su dimensión así como con el total de la escala. Por otro se ha calculado el Alpha de Cronbach que es, indudablemente, uno de los indicadores más potentes para medir la consistencia interna de un instrumento. Y por último y mediante la contrastación de las medias del grupo de pésima/baja conducta exploratoria y alta/óptima conducta exploratoria, se ha calculado el índice de discriminación de cada ítem.

Tal y como aparece en la tabla 6 el Alpha de Cronbrach obtenido de toda la escala es de 0,8967, coeficiente que confirma que todas las puntuaciones proporcionadas por la escala son fiables. El análisis de la consistencia por dimensiones ofrece unos índices de fiabilidad que oscilan entre 0,7336 (dimensión REACCIONES A LA EXPLORACIÓN) y 0,8841 (dimensión PROCESO DE EXPLORACIÓN). El menor índice de fiabilidad obtenido puede ser debido por el número de ítems que engloba esta dimensión (11 preguntas).

TABLA 6
ÍNDICES DE HOMOGENEIDAD Y FIABILIDAD

				CORRELACIÓN		
		Media	Desv. típica	Subdimensión	Dimensión	TOTAL
PROCESO DE EXPLORACIÓN	Item1	2,73	0,989	0,7765**	0,4906**	0,2980**
	Item2	2,58	0,969	0.7756**	0,5548**	0,3712**
	Item3	3,02	1,162	**0,6519****	**0,2192***	**0,1681**
	Item8	3,13	1,143	0,8156**	0,5607**	0,3254**
	Item9	3,09	1,094	0,8033**	0,5125**	0,4460**
	Item1	3,42	1,059	0,8462**	0,5213**	0,3605**
	Item1	2,82	1,338	0,8884**	0,5729**	0,3385**
	Item1	2,77	1,098	0,8284**	0,6972**	0,4573**
	Item1	2,92	1,348	0,7300**	0,5756**	0,2884**
	Item1	2,65	1,240	0,7344**	0,6810**	0,4503**
	Item1	1,98	1,220	0,7335**	0,4587**	0,1842
	Item1	2,47	1,305	0,8042**	0,6328**	0,4314**
	Item1	2,62	1,266	0,6950**	0,5336**	0,3269**
	Item1	2,76	1,668	0,7464**	0,4711**	0,3173**
	Item1	2,55	1,103	0,7173**	0,6907**	0,4267**
	Item2	3,23	1,103	**0,7271****	**0,3109****	**0.1708**
	Item2	3,62	1,068	0,7651**	0,3780**	0,3057**
	Item2	3,23	1,282	0,7813**	0,2846**	0,2682**
	Item2	3,78	1,077	0,7946**	0,4021**	0,4027**
	Item2	3,30	1,205	0,7523**	0,5568**	0,4232**
	Item2	3,00	1,205	1.000	0,6145**	0,5253**
	Item2	1,67	0,881	1.000	0,6410**	0,4248**
REACCIONES A LA EXPLORACIÓN	Item4	2,86	0,901	0,7334**	0,3676**	0,3994**
	Item5	2,99	0,920	0,8442**	0,2907**	0,3269**
	Item6	2,62	1,014	0,6993**	0,3406**	0,3952**
	Item7	3,31	1,043	0,7065**	0,2799**	0,4197**
	Item4	3,30	1,157	0,7282**	0,6264**	0,4527**
	Item4	3,52	1,230	0,7597**	0,4798**	0,3327**
	Item4	3,82	1,192	0,7385**	0,6192**	0,4142**
	Item4	3,50	1,346	0,8142**	0,6415**	0,3389**
	Item4	3,54	1,226	0,8183**	0,6830**	0,4144**
	Item4	3,41	1,203	0,7484**	0,6032**	0,4424**
	Item5	3,40	1,289	0,6879**	0,6546**	0,5354**
EXPECTATIVAS	Item26	3,26	1,135	0,8301**	0,4135**	0,4120**
	Item27	3,14	1,053	0,9113**	0,4677**	0,5026**
	Item28	2,88	1,154	0,7995**	0,3554**	0,3565**
	Item30	3,62	1,175	0,8156**	0,5810**	0,4710**
	Item31	3,78	1,097	0,8366**	0,6266**	0,5107**
	Item32	3,62	1,091	0,7354**	0,5200**	0,3755**
	Item33	3,67	1,162	0,7259**	0,6855**	0,6319**
	Item34	3,46	1,228	0,8745**	0,6510**	0,5971**
	Item35	3,48	1,209	0,8016**	0,7084**	0,5815**
	Item36	3,38	1,210	0,8862**	0,5893**	0,5254**
	Item37	3,97	1,070	0,6965**	0,6141**	0,4712**
	Item38	3,80	1,085	0,7921**	0,4762**	0,2945**
	Item39	3,78	1,097	0,7598**	0,5823**	0,4895**
	Item40	3,77	1,206	0,7086**	0,4465**	0,3250**
	Item41	3,90	1,073	0,7807**	0,4408**	0,2984**
	Item42	3,52	1,161	0,7259**	0,4989**	0,3575**
	Item43	4,33	0,990	0,7361**	0,5624**	0,4400**

ALPHA de CRONBACH = 0,8967
ALPHA de la Dimensión PROCESO DE EXPLORACIÓN = 0,8841
ALPHA de la Dimensión REACCIONES A LA EXPLORACIÓN = 0,7336
ALPHA de la Dimensión EXPECTATIVAS = 0,8516
* significativo con un margen de error 5%, ** significativo con un margen de error 1%

BEST COPY AVAILABLE

Ahora bien, un estudio detallado de la influencia de cada ítem en la fiabilidad total tanto al nivel de toda la escala como al nivel de la dimensión nos dice que:

- La fiabilidad total de la escala aumentaría ligeramente si se eliminaran los ítems 3, 45 y 47 obteniendose un alpha de 0,8977, 0,8974 y 0,8975 respectivamente.
- Que la fiabilidad de la dimensión de PROCESO DE EXPLORACIÓN aumentaría si se eliminase el ítem 3 presentando un alpha de 0,8893.
- En la dimensión REACCIONES A LA EXPLORACIÓN si eliminásemos el ítem 7 se apreciarían algunas diferencias en la fiabilidad aumentando a un 0,7407.
- Y en la última dimensión EXPECTATIVAS la fiabilidad aumentaría su valor si no apareciese la pregunta 28 (0,8559).

Otra forma de estudiar la consistencia interna es mediante la homogeneidad de sus ítems, es decir, el hecho de que cada uno de los ítems mida una parte proporcional de lo que mide la prueba en su conjunto.

En la tabla 6 aparecen los diferentes índices de homogeneidad calculados, mediante la correlación de PEARSON de cada ítem con el total de la escala, el total de la dimensión correspondiente y el total de la subdimensión. Se comprueba que la mayoría de los ítems guardan una correlación significativa con el total, únicamente cabe destacar que en los ítems 3 (*Tengo dificultades/claro para saber en qué trabajaré...*), 15 (*Asistir a algún servicio o programa de orientación profesional*) y *20 (Reflexionar sobre la relación entre mi pasado escolar y mi futuro profesional)* el resultado de la correlación ha sido no significativo, sin embargo, estas mismas preguntas en el estudio de la relación con su dimensión y subdimensión el resultado si ha sido significativo, con un margen de error del 1% y siendo, en todos los casos, el coeficiente de correlación menor de su dimensión.

Son las preguntas «*autoevaluarme para encontrar trabajo*», «*aprender más de mi mismo*» y «*comprender la relación entre mi conducta pasada con la futura*» las que mayor homogeneidad presentan con respecto a la escala total.

4.4. Índice de Discriminación

Este índice nos ayudará a distinguir las preguntas de la escala que discriminan la óptima o pésima conducta exploratoria. Este índice se ha calculado a partir de la aplicación de la prueba paramétrica t student (todos los ítems siguen la Ley Normal), contrastando las medias del grupo con menor puntuación obtenida en la escala con el grupo con mayor puntuación.

El estudio se ha aplicado tanto a las subdimensiones y dimensiones de la escala como a los 50 ítems de la escala. La tabla 7 presenta los resultados obtenidos para cada pregunta, dado que los resultados obtenidos por dimensiones y subdimensiones nos alumbran que todas ellas son significativamente discriminativas, aunque la subdimensión que menor grado de discriminación presenta es la de REACCIONES A LA EXPLORACIÓN.

Tal y como aparece en la tabla 7 la mayoría de los ítems discriminan las conductas exploratorias. De los ítems que han resultado ser no discrimintativos tres corresponden con índices de homogeneidad también bajos (preguntas 3, 15 y 20); los tres

pertenencen a la misma dimensión, PROCESOS DE EXPLORACIÓN aunque en subdimensiones diferentes, **Cantidad de información, Exploración Ambiental y Autoexploración**, destacándose esta última por presentar los índices de discriminación más bajos.

Los ítems 40 y 41 pertenecientes a la subdimensión **Importancia obtener trabajo** dentro de la dimensión EXPECTATIVAS, presentan resultados no significativos y en general se aprecia que la totalidad de ítems que engloban esta subdimensión (40-43) no son muy discriminativos.

TABLA 7
ÍNDICE DE DISCRIMINACIÓN DE LOS ÍTEMS DE LA ESCALA
(SIGNIFICATIVO 5%, ** SIGNIFICATIVO 1%)*

Proceso de Exploración		Reacciones de Exploración		Expectativas	
ITEM	Valor t student	ITEM	Valor t student	ITEM	Valor t student
Item1	4.13**	Item 4	4.29**	Item 26	5.00**
Item 2	4.98**	Item 5	2.62*	Item 27	6.07**
Item 3	1.74	Item 6	4.02**	Item 28	3.63*
Item 8	4.11**	Item 7	4.71**	Item 30	5.02**
Item 9	5.85**	Item 44	4.88**	Item 31	4.90**
Item 10	3.89**	Item 45	2.83*	Item 32	3.75**
Item 11	2.84*	Item 46	3.73**	Item 33	7.08**
Item 12	5.28**	Item 47	3.37*	Item 34	6.43**
Item 13	3.20*	Item 48	4.36**	Item 35	5.89**
Item 14	4.84**	Item 49	4.49**	Item 36	5.92**
Item 15	1.28	Item 50	5.52**	Item 37	3.87**
Item 16	4.01**			Item 38	3.00*
Item 17	3.49*			Item 39	5.15**
Item 18	3.72**			Item 40	1.59
Item 19	4.23**			Item 41	1.58
Item 20	1.41			Item 42	2.71*
Item 21	2.27*			Item 43	3.45*
Item 22	2.03*				
Item 23	3.43*				
Item 24	4.35**				
Item 25	6.37**				
Item 29	4.14**				

Podemos destacar que del resto de ítems que sí son discriminativos, los que mejor discriminan corresponden a las preguntas *«aprender más de mí mismo», «expectativas en*

aquellos trabajos que prefieres en tu área profesional» y *«en estos momentos estoy buscando información principalmente en una ocupación»*.

5. CONCLUSIONES

El análisis estadístico de la escala demuestra la adecuación de la misma para el diagnóstico del constructo de exploración de carrera cuya importancia ha sido documentada en los procesos de transición de los jóvenes. Consideramos que queda justificada la importancia del instrumento para evaluar la conducta exploratoria de adolescentes y jóvenes en proceso de transición al mercado laboral.

El análisis de la escala muestra la importancia de diferenciar las dimensiones cognitiva, conductual y afectiva de los procesos de exploración de la carrera, tanto en contextos diagnósticos como de intervención, campo este último donde estamos asistiendo a importantes avances en el desarrollo de programas. Una revisión actualizada puede verse en el trabajo de Rodríguez Moreno (1999). A partir de una conceptualización del constructo de exploración la autora propone diferentes instrumentos de diagnóstico y programas para su desarrollo en jóvenes.

Sin duda, es necesario continuar trabajando en la escala y realizar otros estudios de validación, sobre todo dirigidos a poblaciones específicas, como jóvenes en vías de exclusión, cuyo proceso de intervención reclama acciones sistemáticas contundentes.

BIBLIOGRAFÍA

ALTMAN, J.M. y SEDLACEK, W.E. (1991). Differences in volunteer interest by level of career orientation. *Journal of Employment Counseling, 28,* 121-128.

BENOIT-GUILBOT, O. (1990). La recherche d'emploi: stratégies, qualification scolaire ou professionnelle et «qualification sociale». *Sociologie du Travail, 4,* 491-506.

BERLYNE, D.E. (1960). *Conflict, arousal and curiosity.* Nueva York: McGraw-Hill Book Company.

BISQUERRA, R. (1989). Introducción conceptual de análisis multivariable. Barcelona: PPU.

BLANCH, J.M. (1990). *Del viejo al nuevo paro. Un análisis psicológico y social.* Barcelona: PPU.

BLUSTEIN, D.L. (1988). The relationship between motivational processes and career exploration. *Journal of Vocational Behavior, 32,* 340-357.

BLUSTEIN, D.L. (1989). The role of goal instability and career self-efficacy in the career exploration process. *Journal of Vocational Behavior, 35,* 194-203.

BLUSTEIN, D.L. (1992). Applying current theory and research in career exploration to practice. *The Career Development Quaterly, 41,* 174-184.

DONOSO, T.; FIGUERA; P., RODRÍGUEZ, M.L. (1994). Un instrumento para evaluar la conducta exploratoria en el desarrollo de la carrera profesional. *Revista de Investigación Educativa, 23,* 490-496.

FIGUERA, P. (1996). *La inserción del universitario en el mercado de trabajo.* Barcelona: EUB.

HEPPNER, P.P. y COOK, S.W. (1991). An investigation of coping styles and gender differences with farmer in career transition. *Journal of Counseling Psychology, 38(2),* 167-174.

HOYT, K.B. (1995). El concepto de educación para la carrera y sus perspectivas. En RODRÍGUEZ MORENO, M.L. (coord.), *Educación para la carrera y diseño curricular (pp. 15-37).* Barcelona: Publicaciones de la Universidad de Barcelona.

JORDAAN, J.P. (1963). Exploration behavior: The formation of self and occupational concepts. In D.E. SUPER (Ed.), *Career development: Self-concept theory* (pp. 42-78). Nueva York: College Entrance Examination Board.

LENT, L.W., BROWN, S.D. y HACKETT, G. (1994). Toward a unified social cognitive theory of career and academic interest, choice and performance[monograph]. *Journal of Vocational Behavior.* 45, 79-122.

LENT L.W. y HACKETT G. (1994). Sociocognitive mechanisms of personal agency in career development. En M. L. Savickas y R.W. Lent (eds.), *Convergence in career development theories: Implications for science and practice* (pp. 77-101). Palo Alto, CA: Consulting Psichologists Press.

LENT, L.W. y BROWN, S.D. (1996a). Applying social cognitive theory to career counseling. *The Career Development Quaterly, 44(4),* 307-309.

LENT, L.W. y BROWN, S.D. (1996b). Social cognitive approach to career development. *The Career Development Quaterly, 44(4),* 310-321.

LENT, L.W. y WORTHINGTON, R.G. (1999). A Social Cognitive View of School-to-Work Transition Process. *The Career Development Quaterly, 47(4),* 297-311.

PEREIRA, M. (1995). Las necesidades de educación para la carrera de estudiantes asturianos al final de secundaria. Bordón, 47(1), 67-77.

RISK, J.W. (1987). The recruitment process for school-leavers: Practical and teoretical implications. *British Journal of Guidance and Counseling, 13(3),* 297-312.

RODRÍGUEZ MORENO, M.L. (coord.)(1995). *Educación para la carrera y diseño curricular.* Barcelona: Publicaciones de la Universidad de Barcelona.

RODRÍGUEZ, M.L., SANDÍN, M.P. y BUISAN, C. La conducta exploratoria: concepto y aplicaciones. *Revista de Educación* (en prensa).

RODRÍGUEZ MORENO, M.L. (1999). *Enseñar a explorar el mundo del trabajo.* Málaga: Ediciones Aljibe.

STEFFY, B.D., SHAW, K.N. y NOE, A.W. (1989). Antecedents and consequences of job search behaviors. *Journal of Vocational Behavior, 35(3),* 254-269.

STUMPF, S.A., COLARELLI, S.M. y HARTMAN, K. (1983). Development of the Career Exploratory Survey (CEX). *Journal of Vocational Behavior, 22,* 191-227.

STUMPF, S.A., ASTIN, E.J. y HARTMAN, K. (1984). The impact of career exploration and interview readiness on interview performance and outcomes. *Journal of Vocational Behavior, 24,* 221-235.

STUMPF, S.A., BRIEF, A.P. y HARTMAN, K. (1987). Self-efficacy expectations and coping with career-related events. *Journal of Vocational Behavior, 31,* 91-108.

STUMPF, S.A. y LOCKHART, M.C. (1987). Career Exploration: Work-rol salience, work preferences, belief and behavior. *Journal of Vocational Behavior, 30,* 258-269.

SUPER, D.E. (1994). A life-span, life-space perspective on convergence. En M.L. SAVICKAS y R.L. LENT (EDS.), *Convergence in career development theories* (pp. 63-74). San Francisco: Jossey-Bass.

TAVEIRA, M.C., SILVA, M.C., RODRÍGUEZ, M.L. (1998). Individual characteristics and career exploration in adolescence. *British Journal of Guidance & Counselin, 26(1),* 89-104.

TAYLOR, M.S. (1985). The roles of occupational knowledge and vocational self-concept crystallization in students' school-to-work transition *Journal of Counseling Psychology, 32(4),* 539-550.

VISAUTA, B. (1998). Análisis estadístico con SPSS para Windows. Madrid: McGrawHill.

TRABAJOS METODOLÓGICOS

Revista de Investigación Educativa, 2000, Vol. 18, n.º 1, págs. 223-242

CRITERIOS DE VALIDEZ EN LA INVESTIGACIÓN CUALITATIVA: DE LA OBJETIVIDAD A LA SOLIDARIDAD

M. Paz Sandín Esteban[*]
Universidad de Barcelona

RESUMEN

En este artículo se realiza una aproximación a las diversas posturas acerca de los criterios de rigor científico en la investigación cualitativa. Para ello, se presenta una revisión del significado que hoy en día ha adquirido el concepto de «validez» en el marco de las ciencias sociales enfatizando su re-elaboración en términos de construcción social del conocimiento. Asimismo se presenta un recorrido desde los enfoques pos-positivistas hasta el pensamiento posestructural en relación a la credibilidad de los resultados obtenidos en los estudios cualitativos. También, se recogen algunos estándares de validez aplicables tanto a la investigación cuantitativa como cualitativa y los procedimientos más utilizados en la práctica investigadora para garantizar los criterios de validez.

ABSTRACT

In this paper it is carried out an approach to the diverse postures about the scientific rigor in qualitative research. It exposes a revision of the meaning that the concept of «validity» has acquired lately in social sciences emphasizing its re-elaboration in terms of social construction of the knowledge. Also it is presented a view in relation to the credibility of the results obtained

* Mª Paz Sandín Esteban. Campus Vall d'Hebron. Passeig Vall d'Hebron, 171. Facultat de Pedagogia. Edifici Llevant. Dept. MIDE. 08035 Barcelona. E-mail: mimse01d@d5.ub.es

in qualitative studies from postpositivism until poststructuralism thought. Also, some standards of validity applicable so much to the quantitative investigation as qualitative are showed and the procedures more used in the research practice to guarantee validity criteria.

1. INTRODUCCIÓN

Los criterios de rigor científico en la investigación cualitativa han ido evolucionando desde propuestas vinculadas a planteamientos positivistas, pasando por el establecimiento de criterios propios hasta su reconceptualización desde el pensamiento posestructural. El objetivo de este trabajo es presentar una panorámica general de las diversas posturas y subrayar los aspectos más relevantes que en la actualidad emergen en relación a la validez de los trabajos desarrollados desde una perspectiva cualitativa. Como veremos, la complejidad de la validación de los estudios cualitativos requiere superar la idolatría del método pues cuestiones ideológicas, políticas (Orden y Mafokozi, 1999) y morales impregnan su desarrollo. No es posible separar los «objetos» de la investigación de la estructura (contenido, forma y desarrollo) de la misma: la *relacionalidad* del proceso y cuestiones *éticas* impregnan el quehacer de los investigadores educativos. El concepto de *solidaridad*, el conocimiento entendido en términos prácticos y morales, frente al de objetividad, regula la actividad investigadora (Rorty, 1996; Smith, 1992).

2. LA LEGITIMIDAD DE LA INVESTIGACIÓN CUALITATIVA

Aunque la investigación cualitativa posee una dilatada historia en el seno de las disciplinas sociales, en las dos últimas décadas ha tenido que luchar, no tanto, por abrir espacios de indagación y reflexión que admitieran otras formas de acercamiento a la realidad, cuanto por el reconocimiento de su legitimidad. En la actualidad los discursos explícitos sobre cuestiones epistemológicas y metodológicas relacionadas con la investigación están alejados de la defensa de la primacía positivista y parecen discurrir por vías de reconocimiento, reconciliación e integración de métodos (Bericat, 1998; Dendaluce, 1995, 1998). Sin embargo, la cuestión aún no resuelta de cómo valorar la credibilidad de los estudios desarrollados en un marco interpretativo oscurece sus aportaciones. Howe y Eisenhart (1993: 174) describen acertadamente esta situación: «inicialmente, el debate sobre la legitimidad de la investigación cualitativa en educación se estableció en términos de una elección decidida, aunque vagamente descrita, entre una metodología bien pertrechada y un recién llegado sumamente sospechoso».

Desde que irrumpió en el campo de la investigación, el «recién llegado» ha desarrollado una identidad propia. Al tiempo que maduraba y reflexionaba sobre cuestiones ontológicas, epistemológicas y metodológicas ha ido adaptando, generando y revisando diversas propuestas sobre los criterios más adecuados para valorar sus aportaciones (Angulo, 1990; Cajide, 1992). Los criterios de validez de la investigación social, y en particular de la investigación cualitativa, han estado y siguen estando en

constante revisión: «al intentar reconstruir una realidad, desentrañar sus redes de significado y, en definitiva, comprenderla en toda su profundidad, nos hacen falta indicadores de credibilidad: «¿Es esa la realidad? ¿Está quizás deformada por los instrumentos que se han empleado, por la prisa con que se ha trabajado, por la subjetividad de los informantes, por la arbitrariedad de la información...» (Santos, 1990: 162).

En realidad los investigadores del ámbito educativo debemos aceptar convivir con una «lesser form of knowledge» (Labaree, 1998) que supone determinados beneficios y también limitaciones y que nos lleva a reconstruir constantemente las bases de la disciplina al tiempo que reinterpretar las cuestiones más fundamentales de nuestro campo. Particularmente, la investigación cualitativa en el posmodernismo afronta una doble crisis de representación y legitimación (Denzin, 1994). La primera hace referencia a la imposibilidad de aprehender directamente la realidad. Ésta se construye, primero, a través de la aplicación de nuestros propios esquemas sobre ella y, después, se re-construye en los textos producidos. En el pos-estructuralismo, el texto, el lenguaje, la narración, cobra un papel fundamental (Richardson, 1994). La crisis de legitimación supone una mirada crítica hacia la utilidad de los tradicionales criterios de rigor científico en el contexto de la investigación cualitativa.

Ciertamente, la palabra clave en todo este entramado es la de *interpretación*. Los investigadores cualitativos abordan, fundamentalmente, cuestiones relacionadas con acciones y no con actos. Justamente, ese es el reto que constituye nuestro trabajo, de hecho, el foco de la investigación social y educativa, la interpretación del significado de acciones humanas y sociales. Pero, ¿dónde reside el «significado»? La pregunta nos dirigiría a los diversos enfoques hermenéuticos (hermenéutica de validación u objetiva, crítica y filosófica) pues «cada posición tiene implicaciones para la comprensión de lo que los investigadores cualitativos hacen cuando investigan y para el «status» de las interpretaciones ofrecidas como resultado de sus estudios» (Smith, 1993: 190).

3. LOS «SIGNIFICADOS» DEL CONCEPTO DE VALIDEZ

Cuando hablamos de la calidad de un estudio solemos referirnos a su rigor científico, fiabilidad, veracidad, confiabilidad, plausibilidad, adecuación metodológica, credibilidad, congruencia, etc. Pero, quizá, el término más utilizado sea el de «validez». Una investigación no válida, no es verdadera. Una investigación no válida, no es una buena investigación, no tiene credibilidad. Si los estudios no pueden ofrecer resultados válidos, entonces, las decisiones políticas, educativas, curriculares, etc. no pueden basarse en ellos. La validez ha constituido siempre una preocupación en la investigación educativa; las cuestiones sobre la validez han emergido históricamente en el contexto de la investigación experimental y ahí por tanto surgieron las primeras respuestas. Es un término que aunque reelaborado desde diversas perspectivas se ha mantenido como elemento fundamental para la valoración de la calidad o rigor científico de los estudios en las ciencias sociales a pesar de que algunos autores cuestionan su protagonismo en los estudios cualitativos (Wolcott, 1990). La validez, como sinónimo de verdad, como verdad construida, como verdad interpretada, consensuada,

significue lo que signifique, se convierte en la línea divisoria, el criterio límite que establece la legitimidad, aceptación o confiabilidad de los trabajos de investigación (Scheurich, 1996). En el amplio ámbito de la investigación cualitativa el significado tradicional del concepto de validez ha sido reformulado, fundamentalmente, en términos de *construcción social del conocimiento* otorgando un nuevo énfasis a la *interpretación*.

Mishler (1990: 419) señala que en la investigación «inquiry-guided» el concepto de *validación* es más relevante que el de validez: «la validación es el proceso(s) a través del cual realizamos afirmaciones y evaluamos la credibilidad de observaciones, interpretaciones y generalizaciones. El criterio esencial para dichas valoraciones es el grado en que podemos basarnos en los conceptos, métodos e inferencias de un estudio como base para nuestra propia teorización e investigación empírica». Con esta reformulación la cuestión esencial de validez se convierte en si la comunidad científica evalúa los resultados como suficientemente confiables para basar en ellos su trabajo aspecto que también reconoce Cutcliffe (1999). En la propuesta de Mishler no se acepta el establecimiento de un grupo de *criterios estándar* para valorar la investigación, por el contrario, se sugiere la configuración de una colección de ejemplos relevantes que muestren cómo deben realizarse los estudios.

Según Maxwell (1992) la postura de Mishler es representativa de los enfoques de validez basados en ejemplos (modelo sintagmático de validez) frente a aquéllos basados en tipologías (modelo paradigmático de validez). Maxwell es partidario de la complementariedad de ambas perspectivas y señala que la *comprensión* es un concepto mucho más importante que el de validez. Así, nos presenta cinco tipos de validez relacionadas con las formas de comprensión inherentes a la investigación cualitativa: la *validez descriptiva* hace referencia a la precisión o exactitud con que los hechos son recogidos en los textos o informes sin ser distorsionados por el propio investigador. Además de proporcionar una descripción válida de objetos, acontecimientos y conductas, el investigador trata de comprender, *validez interpretativa*, qué significado tienen para las personas esos objetos, acontecimientos y conductas. La *validez teórica* se relaciona con las construcciones teóricas que el investigador aporta o desarrolla durante el estudio; nos sitúamos en la explicación, más allá de la descripción y la interpretación, en la validez de un informe como *teoría* de un fenómeno. También nos remite a cuestiones relacionadas con la *generalización*, particularmente, hace referencia a la posibilidad de *generalización interna* dentro de una comunidad, grupo o institución hacia personas, acontecimientos y lugares que no han sido directamente observados o entrevistados. Por otra parte, la *generalización externa* se relaciona con la transferencia de los resultados a otras comunidades, grupos o instituciones. Finalmente, la *validez evaluativa* conlleva reconocer y considerar los marcos evaluativos a través de los cuales se atribuye un significado a los hechos observados.

Un aspecto importante de las ideas de Maxwell es que la validez se relaciona fundamentalmente con los textos y no con los datos o los métodos. La validez es relativa a propósitos y circunstancias. De hecho tal y como se verá a lo largo de esta exposición, en el marco del pensamiento pos-estructuralista, el protagonismo que

223

adquieren los relatos de investigación cualitativa traslada el énfasis del establecimiento de un isomorfismo con la realidad estudiada hacia el nivel de suficiencia de evidencia que aportan los textos en relación a las afirmaciones que en ellos se realizan. La narración en sí misma es objeto de valoración y así aparecen voces sobre, por ejemplo, la *calidad de la transcripción* de datos como un aspecto del rigor en la investigación cualitativa (Poland, 1995), cuáles son los *procedimientos* más adecuados para evaluar e interpretar documentos y textos (Manning y Cullum-Swan, 1994), la consideración de su elaboración como un *método de investigación* en sí mismo (Richardson, 1994) y discusiones acerca de *formas de representación y comunicación* alternativas en la investigación (Eisner, 1997; Sandelowski, 1998).

Impregnados por los postulados pos-estructuralistas actuales Altheide y Johnson (1994: 485) presentan así la cuestión acerca de la validez: «¿Cómo deberían las metodologías interpretativas ser juzgadas por aquellos que comparten la idea de que la perspectiva sobre *cómo* el conocimiento es adquirido, organizado e interpretado está relacionado con *qué* afirmaciones se realizan?» Tras analizar diversas posturas actuales acerca de la validez y cuestiones afines los autores sostienen que los principales significados otorgados al término son los siguientes: *validez como cultura*: se reconoce que el investigador refleja, impone, reproduce, escribe y lee desde su propio bagaje cultural; *validez como ideología*: similar al anterior, incide fundamentalmente en aspectos culturales específicos relacionados con el poder social, la legitimación y cuestiones relacionadas con la estructura social; *validez como género*: remite a las asunciones subyacentes al proceso de recogida de datos y de elaboración conceptual relacionadas con cuestiones acerca del poder y la dominación en la interacción social; *validez como lenguaje/texto*: reconoce cómo categorías culturales y visión del mundo particulares que están presentes en el lenguaje y los discursos conducen, limitan, una determinada construcción de la realidad; *validez como aplicabilidad/ayuda*: subraya la utilidad y «empowerment» de la investigación para beneficiar grupos desfavorecidos; *validez como estándares*: las afirmaciones sobre la verdad son múltiples y, por tanto, debemos evadir formas autoritarias de legitimación. En resumen, estos autores sostienen que es posible elaborar informes claros y coherentes que pueden estar abiertos a la ambigüedad y la incertidumbre. Reconociendo que no se debe privilegiar un formato cultural de comunicación anticipan que próximos debates aparecerán acerca de la utilidad del análisis del discurso, la semiótica, la deconstrucción y otros métodos para analizar símbolos culturales e indicadores de la vida social.

Wolcott (1990) desde una postura más radical y en el marco de la investigación etnográfica rechaza el concepto de validez. Preguntarnos acerca de la validez de un estudio es hacer una pregunta equivocada. Es un término absurdo, dice, puesto que no existe una correcta y única interpretación de la realidad y prefiere hablar simplemente de *comprensión*: «el poder de hacer inteligible la experiencia aplicando conceptos y categorías» (p. 146). Aunque no descarta la necesidad del establecimiento de procedimientos rigurosos en el desarrollo de los estudios (de hecho presenta una serie de estrategias procedimentales a considerar) sugiere que una preocupación excesiva por cuestiones de validez es más una distracción que una ayuda y que, en cualquier caso, «cuán válido es ¿suficientemente válido?» (p. 149).

Algunos autores (Cherryholmes, 1988) abordan el significado que se otorga a la *validez de constructo*, aquélla que sustenta el tópico a investigar, desde diversas perspectivas (fenomenología, teoría crítica, investigación etnográfica, análisis pos-estructuralistas) y sostienen que este tipo de validez de carácter discursivo es central en la investigación pues las actividades de los investigadores tienen lugar en el contexto de una ideología e intereses determinados. Así, por ejemplo, en la perspectiva pos-estructuralista se suplanta el término validez por el de autoridad y las formas de legitimación adoptan una visión política (Denzin, 1994): «el pensamiento pos-estructural se asienta sobre la convicción de que se debe separar cualquier texto de sus pretensiones externas hacia la autoridad. Cada texto debe ser aceptado en sus propios términos. Se renuncia al deseo de producir un texto válido y autorizado. Si la validez desaparece, entonces los valores y la política, no una epistemología objetiva, gobiernan la ciencia» (p. 298).

Scheurich (1996) en un interesante artículo en el que realiza una deconstrucción de los procesos a través de los cuales se han ido construyendo los diversos discursos sobre la validez (pos-positivismo, feminismo...) sostiene la hipótesis de que las «numerosas y aparentes construcciones distintas de la validez son simplemente máscaras que esconden una similitud subyacente, un propósito o función que transciende las supuestas diferencias inconmensurables o límites que separan las distintas epistemologías de investigación» (p. 49). El objetivo fundamental de su trabajo es mostrar esta similitud que en esencia remite a la función de poder de la validez, a una práctica política que construye una línea divisoria ideológica, un criterio límite que establece la legitimidad o aceptación de lo verdadero. Y su propuesta nos remite a un desafío que consiste en identificar formas de validez que no construyan al «otro» a través de esquemas propios.

4. CRITERIOS DE VALIDEZ EN LA INVESTIGACIÓN CUALITATIVA

Fundamentalmente, podríamos señalar cuatro posiciones ante el establecimiento y uso de criterios para evaluar la investigación cualitativa:

a) Aplicación de criterios convencionales
b) Aplicación de criterios paralelos o cuasi-fundacionales
c) Aplicación de criterios propios
d) Nuevos criterios para un mundo pos-moderno

a) Aplicación de criterios convencionales

Los criterios tradicionales para valorar la adecuación y calidad de la investigación son los ya conocidos de validez interna, validez externa, fiabilidad y objetividad. Algunos autores sostienen que los estudios cualitativos deben ser juzgados usando los mismos criterios que aquéllos utilizados en la investigación convencional. Uno de los primeros intentos de importar y aplicar la terminología positivista al ámbito de la investigación cualitativa lo encontramos en Kirk y Miller (1986) en su obra *Reliability and validity in qualitative research.* A esta temática estuvo dedicado el primer volumen

de la serie sobre «Métodos de investigación cualitativa» publicado por la prestigiosa editorial Sage lo que pone de manifiesto la preocupación y deseo de la comunidad de investigadores cualitativos por defender y asegurar la legitimidad de sus aportaciones. Otro ejemplo ampliamente difundido en la literatura lo constituye el trabajo de Goetz y LeCompte (1988) quienes desarrollan los criterios antes mencionados ajustándolos al método etnográfico. Sin embargo, cabe señalar que estas autoras en los últimos años, aunque continúan utilizando la misma terminología que en la obra señalada, se han ido alejando de una visión excesivamente apegada a los patrones positivistas reconociendo, entre otros aspectos, el papel fundamental de las distintas audiencias en el proceso de desarrollo y evaluación de las investigaciones (LeCompte y Preissle, 1993).

Desde esta postura podría defenderse que debe dirigirse la atención hacia las particularidades específicas de las distintas metodologías de investigación más que hacia cuestiones de tipo epistemológico pudiéndose elaborar unos criterios generales de valoración para las investigaciones en las ciencias sociales (Howe y Eisenhart, 1993; Swanborn, 1996). Otros autores, por el contrario, sostienen que no es que constituya una postura errónea utilizar los criterios tradicionales de rigor científico para valorar la investigación de tipo cualitativo sino que incluso en el marco de la investigación cuantitativa son inadecuados (Hammersley, 1992).

b) Aplicación de criterios paralelos o cuasi-fundacionales

Se defiende que los estudios deben ser valorados a partir de criterios generados por y desde dentro del propio paradigma cualitativo pues no es legítimo articularlos contra un telón de fondo epistemológico positivista. Se pretende pues a la luz de la tesis kuhniana sobre la inconmensurabilidad paradigmática generar normas propias que se adecúen a los postulados esenciales que regulan el enfoque cualitativo de investigación. Éste sostiene concepciones alternativas acerca de qué es la realidad, cómo se puede conocer, cuál es el objetivo de la investigación social, por qué y para qué investigar y, por tanto, deben elaborarse estándares de calidad propios. Este argumento va tomando cuerpo y fuerza a partir de que «el intruso» que señalaban Eisenhart y Howe se instala con firmeza en el ámbito de la investigación social desarrollando progresivamente su identidad: se asume una ontología relativista, existen múltiples realidades, una epistemología subjetivista, comprensión *vs.* explicación, y un conjunto de procedimientos metodológicos naturalistas.

El trabajo más representativo de esta postura, ya conocido, es la obra de Guba y Lincoln (Guba, 1989; Lincoln y Guba, 1985) en la que los autores proponen cuatro *términos alternativos* para valorar los procesos de investigación desarrollados en el marco de lo que ellos denominan investigación naturalista. Para referirse de forma general a la calidad de la investigación utilizan el término «trustworthiness»; la cuestión clave es la reconstrucción de las perspectivas de las personas estudiadas y la demostración de que la interpretación del investigador es creíble para aquéllas. Durante mucho tiempo el trabajo de Guba y Lincoln ha constituido el referente esencial para valorar los estudios interpretativos. Basándose fundamentalmente en su aporta-

ción, otros autores como Miles y Huberman (1994) han desarrollado propuestas similares al presentar un patrón alternativo a los criterios de rigor convencionales.

TÉRMINOS CONVENCIONALES Y ALTERNATIVOS DE CRITERIOS DE CALIDAD EN LA INVESTIGACIÓN CUALITATIVA

Aspecto	Término convencional	Guba y Lincoln (1985) Guba (1989)	Miles y Huberman (1994)
Valor de verdad	Validez interna	Credibilidad	Autenticidad
Aplicabilidad	Validez externa/Generalización	Transferibilidad	«Fittingness»
Consistencia	Fiabilidad	Dependencia	«Auditability»
Neutralidad	Objetividad	Confirmabilidad	Confirmabilidad

Sin embargo, estas posturas han sido cuestionadas pues se argumenta que simplemente se han trasladado o transportado las preocupaciones, regulaciones o afirmaciones de verdad de la ciencia positivista al ámbito de la investigación naturalista. Así, los esquemas propuestos se mantienen conceptualmente muy cercanos a los criterios convencionales al tiempo que se intenta ajustar o substituir la ontología y epistemología positivistas por la constructivista. Scheurich (1996) alude a esta situación reconociendo en la diáspora pos-positivista a un emigrante que ha llevado consigo la cultura de su lugar de nacimiento, el enfoque convencional, teniendo como resultado una mezcla de viejas y nuevas metáforas.

c) Aplicación de criterios propios

Como hemos visto, el establecimiento de términos paralelos no ha logrado alcanzar el objetivo de articular los supuestos del enfoque cualitativo de investigación y los criterios de validación propios. De hecho, Guba y Lincoln, en posteriores trabajos, han reconocido que los criterios por ellos propuestos son al fin y al cabo paralelos a los convencionales pues tienen sus orígenes y raíces en asunciones positivistas. Y, proponen cinco nuevos criterios para la valoración de lo que ellos denominan la «Fourth Generation Evaluation», los *criterios de autenticidad*, que emanan directamente de los supuestos constructivistas y que, según los autores, podrían haber sido sugeridos por una persona que nunca hubiera oído hablar del positivismo o de sus afirmaciones de rigor (Guba y Lincoln, 1990; Lincoln, 1995). Los criterios de autenticidad propuestos son: *equidad o justicia («fairness»)*, presentación de los puntos de vista sostenidos por las diversas personas; la *autenticidad ontológica* se refiere al grado en que las construcciones émicas, la conciencia acerca de determinadas cuestiones de las personas son mejoradas, maduradas, y reelaboradas en la evaluación; la *autenticidad educativa* supone ir más allá de las propias elaboraciones y contrastarlas con otras diferentes para conocer cómo sistemas de valores alternativos evocan soluciones diversas; la *autenticidad catalítica* se define como el grado en que una acción es estimulada y facilitada por

los procesos de evaluación y, por último, la *autenticidad táctica* sugiere el «empowerment» de los participantes hacia la acción.

En el ámbito específico de la investigación etnográfica Hammersley (1992) describe las propuestas de varios autores y sintetiza sus aportaciones en los siguientes criterios (p. 64):

a) El grado en el que se produce teoría formal/genérica
b) El grado de desarrollo de la teoría
c) La novedad de las afirmaciones
d) La consistencia entre afirmaciones y observaciones empíricas e inclusión de ejemplos representativos de éstas en el informe
e) La credibilidad del informe para los lectores y/o para las personas investigadas
f) El grado en que los resultados son transferibles a otros contextos
g) La reflexividad del informe: el grado en el que se valoran los efectos del investigador y de las estrategias de recogida de información sobre los resultados y/o la cantidad de información sobre el proceso de investigación que se proporciona a los lectores.

Para Hammersley uno de los errores más comunes en cuanto al establecimiento de estándares para valorar los resultados de la investigación social es que no se realiza una clara distinción entre criterios y los medios o evidencias a través de los cuales se juzga si aquéllos han sido logrados. Según el autor los criterios que deben gobernar en las ciencias sociales son la *validez* o *verdad* y la *relevancia*. Hammersley alude a la verdad, término que según el autor parece haberse convertido en una palabra tabú para muchos científicos sociales, asumiendo una teoría de correspondencia de la verdad que supone una representación selectiva más que una reproducción de la realidad. Reconoce que no es posible conocer con certeza el grado en que un relato es verdadero, por lo tanto, debemos juzgar la validez de las afirmaciones a partir de la adecuación de la *evidencia* presentada para apoyarlas. Para decidir acerca de la *suficiencia de evidencia* debemos considerar tres aspectos: si la exposición es suficientemente *plausible* dado el cuerpo existente de conocimientos; si la afirmación es *central* al argumento presentado por el investigador reclamará más evidencia que si es marginal; el nivel de suficiencia dependerá del *tipo de afirmación* realizada: definición, descripción, explicación, teoría. Por otra parte, cabe preguntarnos por la *relevancia* de una investigación en términos de su relación con algún tópico de interés considerando, además, las audiencias.

d) Nuevos criterios para un mundo pos-moderno

En el enfoque pos-moderno se promulga que el carácter de la investigación cualitativa implica que no pueden existir criterios específicos para juzgar sus productos. Dado un mundo de múltiples realidades construidas, ¿cómo es posible discernir entre una versión verdadera y otra que no lo es? En un interesante artículo, Smith (1984) desgrana los argumentos de Guba acerca de la idiosincrasia del paradigma naturalista

y va descubriendo las inconsistencias que atraviesan su discurso debido a la aplicación de criterios fundacionales a supuestos antifundacionales. Hammersley (1992) acepta en parte la postura de Smith, específicamente que no es razonable elaborar unos criterios fijos y verdaderos cuya aplicación algorítmica nos diga qué es y qué no es verdadero con absoluta precisión, pero sugiere que la extensión de sus argumentos nos llevaría a un cierto relativismo. Para Hammersley «cualquier criterio debe ser heurístico, su aplicación debe basarse en asunciones tácitas y siempre cuestionables y dicha aplicación por consiguiente debe estar sujeta a una posible discusión» (p. 60).

Desde el pensamiento pos-estructuralista se defiende que deben elaborarse un nuevo grupo de criterios divorciados de las tradiciones positivista y pos-positivista. Esta reivindicación surge, sobre todo, desde «voces» hasta ahora olvidadas o silenciadas que configuran nuevas epistemologías y ofrecen su visión particular acerca de las cuestiones relacionadas con la representación y legitimidad de la investigación cualitativa que hoy, más que nunca, muestra su gran polifonía: feminismo, estudios étnicos, culturales, marxistas... Así, clase, raza, género y etnicidad perfilan el proceso de investigación. Como señalan Denzin y Lincoln (1998) las nuevas sensibilidades posmodernas incorporan al debate sobre la representación y legitimidad de la investigación cualitativa cuestiones claves para evaluar su trabajo como la verosimilitud, emocionalidad, responsabilidad personal, ética del cuidado (Medina, 1999), práctica política, diálogos con los participantes.

En este marco, la investigación científica se enfrenta a tres nuevos compromisos: en primer lugar, a nuevas y emergentes relaciones con los participantes; segundo, a posicionamientos profesionales, personales y políticos hacia el uso de la investigación y su potencialidad para promover la acción y, finalmente, a una visión de la investigación que posibilite y fomente la justicia social, la diversidad y el discurso crítico (Lincoln, 1995). En definitiva, siguiendo a esta autora, cualquier debate sobre estándares significa un cambio radical sobre qué es la investigación, para qué se investiga y quién debería tener acceso a ella.

Lincoln señala que un meta-análisis de los criterios actuales mostraría que prácticamente todos ellos son claramente *relacionales*, esto es, reconocen y validan las relaciones entre el investigador y las personas que participan en la investigación. Y nos presenta lo que ella denomina *criterios emergentes* de calidad acerca de los que realiza algunas matizaciones: a) algunos criterios pueden aplicarse a tipos o clases específicas de investigación. Para algunos estudios pueden poseer una utilidad o aplicabilidad limitada; b) algunos criterios pueden ser aplicables en determinadas etapas de la investigación y ser menos apropiados en otras; c) la mayoría de los criterios son relacionales. El conocimiento válido es una cuestión de *relación* entre el investigador y aquéllo que se desea conocer; d) no existe un límite claro entre las cuestiones de rigor y éticas. Los criterios emergentes propuestos son (Lincoln, 1995):

1. *Estándares para juzgar la calidad establecidos en la comunidad investigadora.* A modo de ejemplo, la autora nos remite a la propuesta de Elliott *et al.* publicada en su versión definitiva en 1999 y sobre la que hablaremos más adelante. Aunque reconoce la aportación, es especialmente crítica respecto a las implicaciones que

puede suponer sobre el establecimiento del «status» social del conocimiento. Según Lincoln estos estándares se han construido sobre prácticas de poder y realizan una función legitimadora de exclusión.

2. *Perspectiva epistemológica.* Existe un grupo de criterios de calidad asociados con cuestiones epistemológicas que suponen el reconocimiento y explicitación por parte del investigador de su postura frente a aquéllas. Desde el pos-estructuctu-ralismo, se reconoce que no es posible representar una única verdad; cualquier texto siempre es parcial e incompleto y se elabora desde una perspectiva social, cultural, histórica, racial y de género particular. Los textos deben mostrar estas posturas de manera que se reconoce que la imparcialidad y objetividad del autor limitan la calidad de la investigación.
3. *La comunidad como árbitro de calidad.* Aparecen un grupo de criterios que ponen de manifiesto la naturaleza comunitaria de la investigación. Si los estudios tienen lugar en y están dirigidos a una comunidad, se reivindica que la investigación debería repercutir en la misma, y no sólo servir a propósitos políticos y de producción de conocimiento, a través del establecimiento de un diálogo basado en el razonamiento moral y consideraciones sobre la práctica.
4. *Voz.* La voz, quien habla, a quién, desde qué propósitos construye una determinada realidad. Múltiples y alternativas voces deben impregnar los textos en la investigación cualitativa. Voces hasta ahora ignoradas, reprimidas o invisibilizadas en discursos pretendidamente científicos y objetivos.
5. *Subjetividad crítica.* Diversos términos se utilizan para referirse a este criterio emergente: subjetividad crítica, subjetividad transformadora y reflexividad crítica. En esencia, se refiere al reconocimiento de que el investigador no es una figura aislada del contexto, grupo, cultura... que intenta comprender y representar sino que forma parte de él.
6. *Reciprocidad.* Constituye uno de los estándares de calidad fundamentales debido a la indisoluble relación entre investigador y personas participantes en la investigación en los trabajos de carácter interpretativo que implica un profundo sentido de confianza, comprensión, acuerdo, sensibilidad.
7. *«Sacredness».* La experiencia humana y el conocimiento se refleja en una profunda preocupación por la dignidad humana, la justicia y el respeto interpersonal. Los investigadores que conciben la ciencia en este sentido se preocupan por las consecuencias sociales de la investigación científica.
8. *Compartir privilegios.* La autora describe dos investigaciones en las que los autores comparten con las personas investigadas los «beneficios» que aporta la realización del estudio; por ejemplo, en un caso, económicos. Constituye éste un criterio sobre el que la comunidad investigadora debe reflexionar. Aún abordadas desde una postura respetuosa, ¿a quién «pertenecen» las vidas sobre las que investigamos?, se pregunta la autora. Es más, imaginemos, continúa, un mundo académico en el que se juzgara nuestro mérito y promoción académica a partir de nuestra implicación con las personas que participan en la investigación más que sobre una supuesta distancia objetiva.

5. ESTÁNDARES DE VALORACIÓN APLICABLES A LA INVESTIGACIÓN CUANTITATIVA Y CUALITATIVA

Howe y Eisenhart (1993) sostienen que la cuestión de los criterios para valorar la investigación debe abordarse desde la «lógica en uso» asociada a diversas metodologías de investigación: «dada la imposibilidad de que existan criterios absolutamente independientes de los juicios, los propósitos y los valores humanos y, del mismo modo, dada la imposibilidad de una unidad monolítica del método científico (la que surgiría de los sueños del positivismo) los criterios deben anclarse dentro del proceso de investigación» (p. 177). Un aspecto crucial de la posición de los autores es su articulación entre estándares de validez generales y específicos de diseño. Según los autores, los estándares de validez cumplen tres funciones principales: a) permiten economía de pensamiento a la hora de diseñar y evaluar estudios educativos; b) proporcionan el punto de partida para la reflexión y mejora de la investigación educativa; c) sirven como vehículo de comunicación dentro y entre las tradiciones de investigación y para orientar a las personas que se inician en ella. Así, proponen cinco criterios generales para valorar la investigación educativa, cuantitativa y cualitativa, y ejemplifican a través de la valoración de un estudio etnográfico cómo los estándares generales deben relacionarse y analizarse a la luz de las asunciones subyacentes y técnicas metodológicas asociadas con un diseño específico.

Punch (1998) presenta una propuesta de criterios evaluativos que aborda desde la asunción de que existen similaridades en la lógica de investigación que subyace a los enfoques cuantitativos y cualitativos sugiriendo la combinación de métodos. Los criterios que el autor propone se construyen sobre el concepto de «disciplined inquiry» tomado de Cronbach y Suppes (1969) y Shulman (1988), y la importancia del ajuste entre las partes que componen un proyecto de investigación subrayando la articulación entre la naturaleza de las cuestiones que orientan el estudio y los métodos y enfoque que lo desarrollan. Aunque los criterios se presentan para la valoración de proyectos de investigación finalizados, con pequeñas modificaciones, pueden aplicarse tanto a propuestas como a informes de investigación.

Elliott, Fischer y Rennie (1999) han presentado también, recientemente, una propuesta de criterios para valorar la posible publicación de estudios desarrollados desde un enfoque cualitativo en psicología y ámbitos afines. Han elaborado una guía aplicable tanto a estudios cuantitativos y cualitativos, y un grupo de criterios específicos para la investigación cualitativa. Según los autores, este trabajo sirve a cuatro funciones: a) contribuir al proceso de legitimación de la investigación cualitativa; b) proporcionar una forma de revisión científicamente válida y apropiada de los manuscritos, artículos, tesinas y tesis cualitativos; c) fomentar un mejor control de la calidad de la investigación cualitativa a través de una auto y hetero supervisión y d) promover nuevos desarrollos en enfoque y método. El trabajo que nos presentan es el resultado de varios años de recopilación, análisis y meta-evaluación de diversas propuestas de validación en el ámbito de la investiga-

ción cualitativa a través del consenso y acuerdo entre diversos autores relevantes en este campo[2].

GUÍA PARA LA PUBLICACIÓN DE ESTUDIOS CUALITATIVOS DE INVESTIGACIÓN EN PSICOLOGÍA Y CAMPOS AFINES (ELLIOTT ET AL., 1999)

A. Directrices de publicación para enfoques cuantitativos y cualitativos de investigación	**B. Directrices de publicación específicas a la investigación cualitativa**
1. Explicitación del objetivo y contexto científico 2. Métodos apropiados 3. Respeto hacia las personas participantes 4. Especificación de los métodos 5. Discusión apropiada 6. Claridad de la presentación 7. Contribución al conocimiento	1. Reconocimiento de la propia perspectiva 2. Descripción de la muestra 3. Construcción a partir de ejemplificaciones 4. Proporcionar pistas de credibilidad 5. Coherencia 6. Desarrollo de tareas generales *vs.* específicas de investigación 7. Resonancia en los lectores (representación adecuada, comprensión)

6. ESTÁNDARES ESPECÍFICOS DE VALIDACIÓN EN LOS DIVERSOS ENFOQUES DE INVESTIGACIÓN

A pesar de los esfuerzos por elaborar criterios de calidad aplicables a ambas metodologías reconocemos con Creswell (1998) la especificidad de criterios de calidad que demandarían las diversas tradiciones de investigación dada la amplitud de procedimientos insertos bajo el amplio paraguas de la investigación cualitativa. De hecho, este aserto se encuentra también en trabajos ya presentados como los de Howe y Eisenhart (1993) o Goetz y LeCompte (1988). Creswell nos presenta un análisis de las propuestas de diversos autores sobre la credibilidad en el contexto de cinco tradiciones de investigación: estudios biográficos, fenomenología, teoría fundamentada, etnografía y estudio de casos.

2 A partir de un análisis cluster de 40 estándares, los autores formularon 11 principios para su uso en la valoración de artículos. Este grupo de criterios fueron presentados en el año 1993 en un seminario de la *Society for Psychotherapy and Research* para su discusión. El resultado del trabajo conjunto realizado fue enviado a 23 de las personas participantes quienes, a su vez, realizaron nuevas modificaciones y aportaciones que fueron introducidas. La nueva propuesta fue enviada a 12 prestigiosos investigadores cualiativos de diversas posiciones teóricas. Además, fue presentada en 1994 en un symposia del encuentro anual de la *American Psychological Association*. El resultado de todo este proceso ha sido la publicación en 1999 de las «Evolving Guidelines for Publication of Qualitative Research Studies in Psychology and Related Fields» (Elliott *et al.*, 1999).

PERSPECTIVAS, TÉRMINOS, PROCEDIMIENTOS Y ESTÁNDARES EN LAS TRADICIONES DE INVESTIGACIÓN CUALITATIVA

Tradición	Autor(es)	Perspectiva	Términos	Procedimientos	Estándares
Biografía	Denzin (1989b)	Posmodernismo	Interpretación	Descripción gruesa Reflexibilidad	No bien detallados
Fenomenología	Dukes (1984)	Diferentes términos	Verificación Reconocimiento espontáneo	Confirmación por otros investigadores Reflexividad	No bien detallados
	Moustakas (1994)	Diferentes términos	Verdad Validez intersubjetiva	Percepciones individuales Comprobación con otros	No bien detallados
Teoría fundamentada	Strauss y Corbin (1990)	Términos paralelos	Verificación/ validación Validez suplementaria	Relacionar datos con categorías Contrastar resultados con la literatura	Ver Strauss y Corbin (1990)
Etnografía	Fetterman (1989)	Términos paralelos	Validez etnográfica	Triangulación de fuentes de información Feedback con los informantes	Ver Spindler y Spindler (1987)
	Hammersley y Atkinson (1995)	Términos paralelos	Triangulación Validación respondente	Triangulación (múltiples formas) Confirmación participantes	Ver Lofland (1974)
	Thomas (1993)	Posmodernismo	Cociente de verdad	Reflexividad Implicaciones sociales	
Estudio de casos	Stake (1995)	Términos paralelos	Validez	Triangulación (múltiples formas) Confirmación participantes	Ver Stake (1995)

Fuente: (Creswell, 1998: 216).

Creswell no incluye la investigación-acción en su clasificación y, sin embargo, «no olvidemos, la mejor manera de conocer la realidad es intentar transformarla» (Bartolomé, 1992: 33). La investigación-acción asume muchos de los postulados del paradigma constructivista pero va más allá de él reconociendo la necesidad de la transformación educativa a través de una indagación autorreflexiva de la propia práctica. Waterman (1998) reconoce la utilidad de las formas de validez asociadas con la investigación cualitativa pero reivindica otras más relevantes y pertinentes a la naturaleza de los

procesos de investigación-acción: la *validez dialéctica* (en relación a tensiones que aparecen al articular la práctica, la investigación y la teoría), la *validez crítica* (responsabilidad moral: intenciones, acciones, implicaciones éticas y consecuencias de acciones y teorías) y la *validez reflexiva* (exploración, reconocimiento y explicitación de la perspectiva del investigador).

7. PROCEDIMIENTOS DE VERIFICACIÓN MÁS UTILIZADOS

Finalmente, tras revisar diversas aportaciones sobre los criterios de validez en la investigación cualitativa, presentamos una serie de procedimientos de verificación que parecen ser los más extendidos en este ámbito. Al fin y al cabo las posiciones, ideas deben ser trasladadas a la práctica: «los principales esfuerzos del investigador se dan en la primera línea de la actividad investigadora en medio de constantes decisiones. Es allí donde las filosofías, teorías y metodologías son aplicadas, probadas, aceptadas, adaptadas o rechazadas» (LeCompte y Preissle, 1993: 316).

A partir de una revisión de los principales criterios de *verificación*, término que utiliza frente al de validez para designar la especificidad y legítima aportación de la investigación cualitativa, Creswell (1998) identifica los más abordados en la literatura. Encontramos en Johnson (1997) aunque con algunas modificaciones terminológicas respecto a la revisión anterior una recopilación similar. En el cuadro siguiente se presentan sus aportaciones.

PROCEDIMIENTOS MÁS UTILIZADOS EN LA OPERATIVIZACIÓN DE CRITERIOS DE RIGOR EN LA INVESTIGACIÓN CUALITATIVA

(Johnson, 1997)	(Creswell, 1998)
1. Investigador como «detective»	1. Trabajo prolongado y observación persistente
2. Trabajo de campo extenso	2. Triangulación
3. Utilización de descriptores de baja inferencia	3. Revisión jueces
4. Triangulación: de datos, de métodos, de jueces, teórica	4. Análisis de casos negativos
5. Validación desde los participantes	5. Clarificación de los sesgos del investigador
6. Revisión jueces	6. Confirmación con los participantes
7. Búsqueda y análisis de casos negativos	7. Descripción gruesa
8. Reflexividad	8. Auditoría externa
9. Establecimiento de patrones	

Vemos que los procedimientos más utilizados son los que básicamente introdujeron Guba y Lincoln en sus primeros trabajos, aunque, una lectura atenta a las aportaciones que en la actualidad se realizan desde la investigación, por ejemplo, feminista y desde una perspectiva posestructural, en general, descubre el intento de generar e introducir nuevas vías de validación y construcción del conocimiento. Desde este enfoque uno de los procedimientos que con más fuerza está apareciendo e inundando

los discursos sobre la validación de los estudios cualitativos es la *reflexividad*. Más allá de su tradicional consideración metodológica, la reflexividad se convierte en una cuestión ideológica acerca de dimensiones éticas, políticas, sociales y personales relacionadas con la investigación y, concretamente, se refiere a la conciencia y autocrítica reflexiva que el investigador realiza sobre sí mismo en relación a sus predisposiciones y los posibles sesgos que pueden afectar al proceso de investigación y los resultados. Marcus (1994), por ejemplo, aborda este tema en el marco de la investigación etnográfica y nos presenta cuatro estilos de reflexividad. Barry *et al.* (1999) sugieren que la reflexividad no es sólo una actividad individual sino una estrategia prioritaria en las investigaciones cualitativas que se desarrollan en grupo porque fomenta la calidad y el rigor de los estudios.

Por el contrario, el tan aclamado procedimiento de *triangulación* se enfrenta a voces críticas que cuestionan su relevancia y posibilidad de garantizar la credibilidad de los resultados. Aunque también se pone en duda la necesidad de la *triangulación de métodos* (Bloor, 1999) quizá es la *validación desde los participantes* la estrategia más discutida (Bloor, 1999; Cutcliffe y McKenna, 1999; Smith, 1984; Swanborn, 1996). En primer lugar, supondría la existencia de fenómenos inalterables. Por otro lado, los investigadores modifican la realidad a validar al facilitar una explicación de la misma pues las personas reinterpretan su pasado y su presente influidos por dicha información. Así mismo, ¿es posible que las personas participantes comprendan los constructos elaborados por los investigadores?; además, no todas ellas pueden ofrecer su opinión: niños/as, personas discapacitadas, que no conocen el idioma, etc.; ¿debe el investigador verificar sus conclusiones con todas y cada una de las personas?, ¿debe intentar contrastar todos los conceptos, categorías o teorías o sólo una parte de ellas? Y, en cualquier caso, finalmente, lo que conseguiríamos es el consenso, no la verdad: «aquí la moraleja es simplemente que es importante llegar a un consenso sobre la obra de uno, pero una vez que lo tenemos, no se debería pensar que hemos acaparado la verdad. Lo que se ha acaparado es el acuerdo» (Eisner, 1998: 76).

8. CONSIDERACIONES FINALES

Como se ha presentando a lo largo de este trabajo, las distintas posturas relativas a la elaboración de conocimiento científico se articulan alrededor de dos cuestiones fundamentales a las que Eisner (1998) denomina, basándose en la distinción realizada por Newell (1986), *objetividad ontológica* y *objetividad de procedimiento*. La primera hace referencia a la voluntad de establecer una versión verdadera, podríamos decir, objetiva, de la realidad: «en el mejor de todos los mundos, buscamos la verificabilidad tanto en la percepción como en el entendimiento. Lo que deseamos ver y saber no es un mundo subjetivo, hecho de creencias a través de la fantasía, la ideología o el deseo, sino lo que está realmente ahí fuera» (p. 59). Por otro lado, los distintos autores proponen una forma particular de objetividad de procedimiento, es decir, «el desarrollo y utilización de un método que elimina, o aspira a eliminar, la incumbencia del juicio personal en la descripción y valoración de un estado de hechos» (p. 60). Eisner propone una alternativa a la dicotomía objetividad/subjetividad y sostiene que la

experiencia humana es una «transacción», es el resultado de la interacción entre dos entidades *postuladas*, lo objetivo y lo subjetivo: «ninguna objetividad prístina ni ninguna subjetividad pura son posibles (...) toda experiencia que se derive del texto es transactiva» (p. 71).

En relación a la generación de conocimiento científico, cabe destacar la importancia que algunos autores otorgan a la aportación de la investigación cualitativa al desarrollo teórico. De hecho, ésta debe afrontar el reto y la responsabilidad de contribuir a la ampliación y desarrollo del *corpus* de conocimiento existente. Por ejemplo, en trabajos como los de Bartolomé y Anguera (1990), en el marco de la investigación cooperativa, y Bartolomé *et al.* (1997), desde un enfoque etnográfico, se muestra que es posible extraer conocimiento a nivel de interpretación y meta-análisis. Desde esta perspectiva, Morse (1997) después de realizar una comparación de las características de la teoría derivada de estudios cuantitativos, cualitativos y prácticos sugiere que, más allá de los tradicionales criterios metodológicos de credibilidad en la investigación cualitativa, debemos extender y ampliar el análisis juzgando *el nivel de teoría que se deriva de los resultados de trabajos cualitativos*. Siguiendo este enfoque, y sin despreciar los trabajos de investigación que se desarrollan en contextos específicos, debemos atender a las tendencias internacionales de investigación (por ejemplo, «*Fifth Framework Programme*, 1998-2002») que apuestan por la realización de estudios comparativos en diversos países, el establecimiento de indicadores comunes y el meta-análisis teórico.

Asimismo, el énfasis que los diversos autores atribuyen a la validez como acción, utilidad o «empowerment» nos recuerda la estrecha relación entre investigación e intervención, entre pensamiento y decisión (Tejedor, 1995). En ese sentido, recordemos la importancia de la *transformación* como criterio de validez en los procesos de investigación-acción así como de la «cristalización del cambio» como indicador del nivel de incidencia e impacto del proceso.

En resumen, la panorámica presentada es una muestra de la densidad de los debates acerca de los criterios de valoración de los estudios cualitativos en el ámbito de las ciencias sociales que pone de manifiesto la complejidad del campo y que nos invita a reflexionar acerca de nuevas responsabilidades personales y profesionales, políticas, éticas y sociales relacionadas con la investigación educativa.

BIBLIOGRAFÍA

ALTHEIDE, D. L. y JOHNSON, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485-499). Londres: Sage.

ANGULO, J. F. (1990). El problema de la credibilidad y el lugar de la triangulación en la investigación interpretativa: un análisis metodológico. En J. B. Martínez Rodríguez (Ed.), *Hacia un enfoque interpretativo de la enseñanza* (pp. 95-110). Granada: Servicio de Publicaciones de la Universidad de Granada.

BARRY, C.A., BRITTEN, N., BARBER, N., BRADLEY, C. y STEVESON, F. (1999). Using reflexivity to optimize teamwork in qualitative research. *Qualitative Health Research, 9*(1), 26-44.

BARTOLOMÉ, M. (1992). Investigación cualitativa en educación: ¿comprender o transformar? *Revista de Investigación Educativa, 20*(2), 7-36.

BARTOLOMÉ, M. y ANGUERA, M. T. (1990). *La investigación cooperativa: vía válida para la innovación en la universidad*. Barcelona: PPU.

BARTOLOMÉ, M. (coord.) *et al.* (1998). *Diagnóstico a la escuela multicultural*. Barcelona: Cedecs.

BERICAT, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social*. Barcelona: Ariel.

BLOOR, M. (1999). On the analysis of observational data: A discussion of the worth and uses of inductive techniques and responden validation. En A. Bryman y R. G. Burgess (Eds.), *Qualitative research. Volume III* (pp. 445-454). Londres: Sage.

CAJIDE, J. (1992). La investigación cualitativa: Tradiciones y perspectivas contemporáneas. *Bordón, 44*(4), 357-373.

CHERRYHOLMES, C.H. (1988). Construct validity and the discourses of research. *American Journal of Education, 96*, 421-457.

CONSTAS, M.A. (1998). The changing nature of educational research and a critique of postmodernism. *Educational Researcher, 27*(2), 26-33.

CRESWELL, J. W. (1998). *Qualitative inquiry and research design. Choosing among five traditions*. California: Sage.

CRONBACH, L. J. y SUPPES, P. (Eds.). (1969). *Research for tomorrow's schools: Disciplined inquiry for education*. Nueva York: Macmillan.

CUTCLIFFE, J. R. y MCKENNA, H. P. (1999). Establishing the credibility of qualitative research findings: the plot thickens. *Journal of Advanced Nursing, 30*(2), 374-380.

DENDALUCE, I. (1995). Avances en los métodos de investigación educativa en la intervención psicopedagógica. *Revista de Investigación Educativa, 26*(2), 9-32.

DENDALUCE, I. (1998). Algunos retos metodológicos. *Revista de Investigación Educativa, 16*(1), 7-24.

DENZIN, N.K. (1994). Evaluating qualitative research in the poststructural moment: the lessons James Joyce teaches us. *Qualitative Studies in Education, 7*(4), 295-308.

DENZIN, N.K. y LINCOLN, Y., S. (1998). Entering the field of qualititative research. En N. K. Denzin y Y. Lincoln, S. (Eds.), *Collecting and interpreting qualitative materials* (pp. 1-34). Londres: Sage.

EISENHART, M. A. y HOWE, K. R. (1992). Validity in educational research. En M. D. LECOMPTE, W. L. MILLROY y J. PREISSLE (Eds.), *The handbook of qualitative research in education* (pp. 642-679). California: Academic Press.

EISNER, E. W. (1997). The promise and perils of alternative forms of data representation. *Educational Researcher, 26*(6), 4-10.

EISNER, E. W. (1998). *El ojo ilustrado. Indagación cualitativa y mejora de la práctica educativa*. Barcelona: Paidós.

EISNER, E. W. y PESHKIN, A. (Eds.). (1990). *Qualitative inquiry in education. The continuing debate*. Nueva York: Teachers College Press.

ELLIOTT, R., FISCHER, C. T. y RENNIE, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology, 38*, 215-229.

GOETZ, J.P. y LECOMPTE, M.D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.

GUBA, E.G. (1989). Criterios de credibilidad en la investigación naturalista. En J. Gimeno Sacristán y A. Pérez Gómez (Eds.), *La enseñanza: su teoría y su práctica* (3ª ed., pp. 148-165). Madrid: Akal.

GUBA, E.G. y LINCOLN, Y.S. (1990). *Fourth generation evaluation*. (2ª ed.). London: Sage.

HAMMERSLEY, M. (1992). *What's wrong with ethnography?* Londres: Routledge.

HOWE, K. y EISENHART, M. (1993). Criterios de investigación cualitativa (y cuantitativa). Prolegómenos. *Revista de Educación, 300*, 173-189.

JOHNSON, R.B. (1997). Examining the validity structure of qualitative research. *Education, 118*(2), 282-292.

KIRK, J. y MILLER, M. L. (1986). *Reliability and validity in qualitative research*. Londres: Sage.

LABAREE, D. F. (1998). Educational researchers: living with a lesser form of knowledge. *Educational Researcher, 27*(8), 4-12.

LECOMPTE, M.D. y PREISSLE, J. (1993). *Ethnography and qualitative design in educational research*. (2 ed.). San Diego, California: Academic Press.

LINCOLN, Y. y GUBA, E. (1985). *Naturalistic inquiry*. Berverly Hills, California: Sage.

LINCOLN, Y.S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry, 1*(3), 275-289.

MANNING, P.K. y CULLUM-SWAN, B. (1994). Narrative, content, and semiotic analysis. En N.K. Denzin y Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 463-477). Londres: Sage.

MARCUS, G.E. (1994). What comes (just) after «post»? The case of ethnography. En N. K. Denzin y Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 563-574). Londres: Sage.

MAXWELL, J.A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279-300.

MEDINA, J.L. (1999). *La pedagogía del cuidado. Saberes y prácticas en la formación universitaria en enfermería*. Barcelona: Laertes.

MILES, M.B. y HUBERMAN, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. Londres: Sage.

MISHLER, E.G. (1990). Validation in inquiry-guided resarch: The role of exemplars in narrative studies. *Harvard Educational Review, 60*(4), 415-443.

MORSE, J.M. (1997). Considering theory derived from qualitative research. En J.M. Morse (Ed.), *Completing a qualitative project. Details and dialogue* (pp. 163-189). California: Sage.

ORDEN HOZ, A. DE LA y MAFOKOZI, J. (1999). La investigación educativa: naturaleza, funciones y ambigüedad de sus relaciones con la práctica y la política educativas. *Revista de Investigación Educativa, 17*, 1, 7-29.

POLAND, B.D. (1995). Transcription quality as an aspect of rigor in qualitative research. *Qualitative Inquiry, 1*(3), 290-310.

PUNCH, K.F. (1998). *Introduction to social research. Quantitative and qualitative approaches*. Londres: Sage.

RICHARDSON, L. (1994). Writing: A method of inquiry. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 516-529). Londres: Sage.

RORTY, R. (1996). *Objetividad, relativismo y verdad*. Barcelona: Paidós.

SANDELOWSKI, M. (1998). Writing a good read: Strategies for re-presenting qualitative data. *Research in Nursing and Health, 21*, 375-382.

SANTOS, M. A. (1990). *Hacer visible lo cotidiano: Teoría y práctica de la evaluación cualitativa de los centros escolares*. Madrid: Akal.

SCHEURICH, J.J. (1996). The masks of validity: a deconstructive investigation. *Qualitative Studies in Education, 9*(1), 49-60.

SHULMAN, L.S. (1988). Disciplines fo inquiry in education: an overwiew. En R.M. Jaeger (Ed.), *Complementary methods for research in education* (pp. 3-17). Washington, DC: American Educational Research Association.

SMITH, J.K. (1984). The problem of criteria for judging interpretive inquiry. *Eductional evaluation and policiy analysis, 6*(4), 379-391.

SMITH, J.K. (1992). Interpretive inquiry: a practical and moral activity. *Theory into Practice, 31*(2), 100-106.

SMITH, J.K. (1993). Hermeneutics and qualitative inquiry. En D.J. Flinders y G.E. Mills (Eds.), *Theory and concepts in qualitative research. Perspectives from the field* (pp. 183-200). Nueva York: Teachers College Press.

SWANBORN, P.G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality and Quantity, 30*, 19-35.

TEJEDOR, F.J. (1995). La investigación educativa y su contribución al desarrollo de la innovación. *Bordón, 47*, 117-193.

WATERMAN, H. (1998). Embracing ambiguities and valuing ourselves: issues of validity in action research. *Journal of Advancing Nursing, 28*(1), 101-105.

WOLCOTT, H.F. (1990). On seeking -and rejecting- validity in qualitative research. En E. W. Eisner y A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152). Nueva York: Teachers College Press.

ELABORACIÓN Y VALIDACIÓN DE UN SISTEMA DE INDICADORES PARA EVALUAR PROGRAMAS DE ORIENTACIÓN EDUCATIVA

CAT[3]	AÑO[3]	CLASIFICACIÓN[3]
D	1999	61059

N.º CITAS[4]	N.º PÁGINAS
	721

Autora: Nuria Manzano Soto
Dirección: C/ Avda. de los Castillos, 30, 7º A, 28925, Alcorcón-Madrid
Director: Ángel Lázaro Martínez
Dpto: Ciencias Humanas y Sociales
Dirección: Edificio Vives, C/ Luis de Ulloa, s/n - 26004- Logroño
Centro: Universidad de la Rioja

Descriptores

Evaluación de programas, Indicadores educativos, Programas de orientación educativa, Validación.

Bibliografía fundamental

BORDERS, L.D. - DRURY, S.N.: *Counseling programs: a guide to evaluation*, Corwin Press, Newbury Park, London, 1992b.

GYSBERS, N.C. - HENDERSON, P.: *Developing and managing your school guidance program* (2ª ed.), ACA, Alexandria, 1994d.

HENDERSON, P. - GYSBERS, N.C.: *Leading and managing your school guidance program staff*. ACA, Alexandria, 1998.

HERNÁNDEZ FERNÁNDEZ, J. - MARTÍNEZ CLARÉS, P.: Propuesta metodológica para evaluar programas de orientación educativa. *Revista electrónica de investigación y evaluación educativa*, 1996, vol. 2, nº 2-3.

JOHNSON, S.K. - WHITFIELD, E.A. (Eds.): *Evaluating guidance programs: A practicioner's guide*, American college testing publications, Iowa City, 1991b.

LÁZARO MARTÍNEZ, A.: Evaluación de la orientación. En: M.A. VERDUGO ALONSO. *Evaluación curricular*, Siglo Veintiuno, Madrid, 1994c.

RODRÍGUEZ ESPINAR, S.: (Coord.) y otros: *Teoría y práctica de la orientación educativa*, PPU, Barcelona, 1993.

SANZ ORO, R.: *Evaluación de programas de orientación educativa* (2ª ed.), Pirámide, Madrid, 1995a.

THE SCOTTISH OFFICE, EDUCATION DEPARTMENT: *Using Performance Indicators in Secondary School Self-Evaluation*. HM Inspectors for Schools. School Development Planning Support Materials, 1992.

WALZ, G.R. - BLEUER, J.C.: *Counselor efficacy: assessing and using counseling outcomes research*, ERIC/ CASS, Michigan, 1993.

Problema de investigación

El propósito de la investigación fue: La elaboración de un sistema de indicadores para la evaluación de las actividades de orientación educativa en sus diferentes formas de organización: Equipos Psicopedagógicos y actividades de Orientación. El problema es: *¿Responde el Sistema de Indicadores al vacío metodológico existente en los Equipos y Departamentos Psicopedagógicos para evaluar su actividad orientadora?* Por tanto, trata de buscar una metodología fiable —Sistema de Indicadores— que permita responder objetiva y eficazmente a esta cuestión. Sin embargo, antes de ofrecer este instrumento para poder aplicarlo, fue necesario ponerlo a prueba, entendiendo este proceso de «ensayo» como una parte inicial de un proceso de evaluación más amplio. No se trata tanto de juzgar el S.I. como de afinar su diseño y estructura, valorar su viabilidad y determinar condiciones de aplicación y metodología de evaluación.

Para dar respuesta a dicho problema, se acotaron los objetivos de la investigación:

1) Elaborar un sistema de indicadores de evaluación que permita conocer y analizar el estado actual de la orientación educativa, y en último término, optimizar la orientación educativa.

2) Validar el S.I. a través de un juicio de expertos.

3) Realizar prueba-piloto del S.I., con el objeto de concretar el S.I. en el contexto peculiar de cada Equipo de Orientación.

BEST COPY AVAILABLE

Muestra y método de muestreo

Para realizar la **validación de contenido** del Sistema de Indicadores elaborado se recogieron los juicios de valor de expertos en el tema, mediante la técnica Delphi. Dicha muestra de expertos fue intencional u opinática, ya que la selección se hace por expertos, según los siguientes criterios que determinaban el perfil de experto: Académico-universitario, Universitario-profesional, profesional. Se seleccionaron 18 expertos divididos en estos tres grupos equilibrados, y que representaran a priori las diferentes posturas sobre el tema: visión académico/ profesional, titularidad del centro (público/privado/concertado), nivel de acción (D.O/EOEP), diferente ubicación geográfica. Por último, los criterios de valoración para cada indicador fueron tres: Claridad (SI/NO), Aceptación y Ponderación (5 4 3 2 1), Medida (CN, CL, MT).

Para realizar la **concreción del S.I.** no hubo muestra, sino que se utilizó el criterio de la disponibilidad y voluntariedad de los EOEP de Madrid. Se reunieron 11 equipos de orientación.

Metodología de trabajo

La metodología para la **elaboración del S.I.** fue: 1) Especificación del concepto, calidad y objetivos de la orientación; 2) Desarrollar un modelo teórico: con variables de contexto, entrada, proceso y producto; así como criterios de relación como funcionalidad, eficiencia, aplicabilidad, productividad, eficacia/efectividad; 3) Identificar una red jerárquica de factores (6), dimensiones (20) y dimensiones (100); 4) Elaborar la estructura del S.I.: tiene cuatro secciones a) Indicadores, b) Recogida de información, c) Valoración/Decisión, d) Plan de acción.

La metodología para la **validación del S.I.** fue: 1) Selección de Expertos; 2) Diseño del cuestionario de validación; 3) Aplicación y recogida de información a expertos; 4) Análisis de información (Delimitación de los puntos de corte); 5) Valoración general del Sistema de Indicadores (Ponderación definitiva de indicadores, Combinación y consenso de la propuesta de expertos).

La metodología para la **concreción del S.I.** fue: 1) Realizar un análisis DAFO del Equipo o D.O.; 2) Detectar las necesidades de orientación y priorizarlas; 3) Establecer las discrepancias entre el presente y futuro en dichas necesidades; 4) Selección de los indicadores útiles, así como los indicadores útiles pero no factibles; 5) Añadir los indicadores nuevos que sean necesarios.

Técnicas de análisis

Las técnicas de análisis utilizadas para los datos de la validación de expertos fueron: 1) Análisis exploratorio: distribución de frecuencias y porcentajes por indicador; dimensión, factor y grupo de expertos; medidas de tendencia central y de dispersión (para ver la variabilidad y amplitud del S.I.) por indicador, dimensión y factor, así como por grupo de expertos; 2) Análisis exploratorio de conglomerados por sujetos (para las dimensiones y los factores), utilizando el método jerárquico aglomerativo (para ver la congruencia y robustez del S.I.).

Se utilizó el paquete SPSS X. Una vez analizados los datos, se modificó el S.I. de acuerdo a los puntos de corte definidos previamente.

Conclusiones

En la validez de contenido realizada, se llega a las siguientes conclusiones:

1. De los 100 indicadores posibles, sólo 16 salen obligatorios (Imprescindibles). Los demás serán opcionales para cada EOEP. Pertenecen a aspectos de entrada que suelen ser parte del trabajo diario del orientador: I.3.4. Constitución del D.O./EOEP; I.3.5. Coordinación del PCO; I.6.8. Actividades; I.6.10. Personal; I.6.11. Recursos necesarios; I.6.12. Tiempo y secuenciación de actividades; I.6.13. Destinatarios; I.6.15. Valoración del PCO; I.7.1. Nº de Personal; I.8.1. Formación inicial del personal; I.9.1. Funciones y roles profesionales; I.9.2. Responsabilidades del orientador y personal de apoyo; I.9.3. Competencias necesarias; I.9.4. Tareas; I.12.1. Adecuación y accesibilidad de espacios; I.13.2. Materiales de diagnóstico.
2. En el **criterio de claridad**, los indicadores más claros pertenecen a los factores: II. Diseño del PCO y IV. Ejecución del PCO. Los menos claros pertenecen a los factores: V. Planificación de la evaluación y VI. Resultados. El grupo de «prácticos» señala mayor claridad que el resto de expertos.
3. En el **criterio de aceptación y ponderación**, existe gran similitud de respuestas en los 3 grupos de expertos, oscilando entre 5 (Imprescindible), 4 (Importante) y 3 (Opcional), aunque dan mayor ponderación el grupo de «académicos-prácticos». Los factores con más peso son: Diseño, Contexto, Implementación. Los que reciben menor peso son: Evaluación, Recursos y Resultados.
4. En el **criterio de medida**, la más valorada en la medida cualitativa y mixta. Existe gran afinidad en las respuestas de los 3 grupos de expertos y similitud de respuestas en todos los indicadores excepto D10 (cuantitativa) correspondiente a Prepuesto.

En la prospectiva, sería interesante: a) Contextualizar el S.I. en varios EOEP y establecer un sistema de relaciones y correspondencias entre indicadores; b) Asignar pesos a cada uno de los indicadores del sistema; y, c) Elaborar unas Guías de recogida de información para cada indicador.

REVISTA INVESTIGACIÓN EDUCATIVA

BOLETÍN SUSCRIPCIÓN

Para subscribirse llene este boletín y devuélvalo a:
REVISTA DE INVESTIGACIÓN EDUCATIVA
Dpto. «Métodos de Investigación y Diagnóstico en Educación»
Facultad de Pedagogía Paseo del Valle de Hebrón, 171, 2ª planta
08035 - BARCELONA (Spain)

Nombre .....

D.N.I. o N.I.F. .....

Dirección .....

Población ..... C.P. .....

País ..... E-mail:..... Teléfono .....

Coste de la inscripción:

- ☐ Individual: 6.000 + I.V.A. ptas.
- ☐ Institucional: 10.000 + I.V.A. ptas.
- ☐ Números sueltos: 2.500 ptas.
- ☐ Indicar n.º deseado: .....
- ☐ Números extras y monográficos: 3.000 ptas.
- ☐ Indicar n.º deseado: .....

SUSCRIPCIÓN INTERNACIONAL

INDIVIDUAL

- ☐ Europa: 6.000+2.000 ptas. gastos de envío (Fecha y Firma)
- ☐ América: 6.000+3.000 ptas. gastos de envío

INSTITUCIONAL

- ☐ Europa: 10.000+2.000 ptas. gastos de envío
- ☐ América: 10.000+3.000 ptas. gastos de envío

BOLETÍN DE DOMICILIACIÓN BANCARIA

Señores,

Les agradeceré que con cargo a mi cuenta/libreta atiendan los recibos que les presentará la *Revista de Investigación Educativa*, como pago de mi subscripción a la misma.

Titular de la cuenta .....

Banco/Caja .....

N.º de cuenta ..... N.º de libreta .....

Agencia .....

Población .....

(Fecha y Firma)

A.I.D.I.P.E.
BOLETÍN ASOCIADO

Para asociarse llene las dos partes de este boletín y devuélvalo a:

A.I.D.I.P.E.
Dpto. «Métodos de Investigación y Diagnóstico en Educación»
Facultad de Pedagogía **Paseo del Valle de Hebrón, 171, 2ª planta**
08035 - BARCELONA (Spain)
Cuota de suscripción anual 6.000 ptas.

DATOS PERSONALES

Nombre y Apellidos .....
D.N.I. o N.I.F. .....
Dirección .....
Población ..... C.P.....
Provincia ..... E-mail:..... Teléfono ( )..... .....
Deseo asociarme desde el día ..... de..... de 20.....
DPTO. TRABAJO. ..... **CENTRO TRABAJO**.....
Situación profesional ..... Dist. Universitario.....

DATOS BANCARIOS

Titular de la cuenta .....
Banco/Caja .....
Domicilio Agencia .....
Población ..... C.P.....

(Firma)

CODIGO CUENTA CLIENTE			
Entidad	**Oficina**	**D.C.**	**Núm. de Cuenta**

A.I.D.I.P.E.
BOLETÍN DOMICILIACIÓN BANCARIA

Señores,
Les agradeceré que con cargo a mi cuenta/libreta atiendan los recibos que les presentará la **Asociación Interuniversitaria de Investigación Pedagógica** como pago de mi cuota de asociado.

Titular de la cuenta .....
Banco/Caja .....
Domicilio Agencia .....
Población ..... C.P.....

(Fecha y Firma)

CODIGO CUENTA CLIENTE			
Entidad	**Oficina**	**D.C.**	**Núm. de Cuenta**

AIDIPE

Asociación Interuniversitaria de Investigación Pedagógica, creada en 1987

FINES DE LA ASOCIACIÓN

a) Fomentar, estimular e impulsar la investigación educativa de carácter empírico en todas sus vertientes y ámbitos de aplicación.
b) Difundir los resultados de esta investigación.
c) Agrupar a todos los que en España cultivan esta área de conocimiento, como docentes o investigadores.
d) Colaborar con las autoridades educativas y de política científica en la elaboración de instrumentos de medida, evaluación y de diseños de investigación tendentes a la mejora del sistema educativo y de desarrollo de los mismos.
e) Favorecer la formación y el perfeccionamiento especializado del personal investigador y de todos aquellos profesionales que intervienen en la investigación educativa.
f) Velar por la calidad científica y profesional de las actividades relacionadas con la investigación educativa.
g) Velar por el respeto y cumplimiento de las normas deontológicas de la investigación científica referida al campo psicopedagógico.

DERECHOS DE LOS SOCIOS

Los socios numerarios y fundadores tendrán los siguientes derechos:

a) Tomar parte en cuantas actividades organice o patrocine la Asociación en cumplimiento de sus fines.
b) Disfrutar de todas las ventajas y beneficios que la Asociación pueda obtener.
c) Participar en las Asambleas con voz y voto.
d) Ser electores y elegibles para los cargos directivos.
e) Recibir información sobre los acuerdos adoptados por los órganos directivos, sobre las actividades técnicas y científicas que la Asociación desarrolle.
f) Hacer sugerencias a los miembros de la Junta Directiva de la Asociación.
g) Impugnar los acuerdos y actuaciones de la Asociación que sean contrarios a los Estatutos, dentro del plazo de cuarenta días y en la forma prevista por las leyes.

Cada socio tiene derecho a recibir la Revista de Investigación Educativa de forma gratuita.

SEMINARIOS

AIDIPE organiza cada 2 años, un Seminario Nacional con un tema central en torno al cual giran las ponencias, comunicaciones, mesas redondas. Los primeros Congresos han sido los de Barcelona (1981), Sitges (1983), Gijón (1985), Santiago de Compostela (1988), Murcia (1990).

PUBLICACIONES

AIDIPE publica las Actas de sus Seminarios, así como la Revista de Investigación Educativa (RIE) con carácter semestral.

Para más información, ver hoja de inscripción adjunta.

NORMAS PARA LAS COLABORACIONES

El objetivo de la Revista de Investigación Educativa es promover el intercambio de información acerca de investigaciones empíricas de carácter educativo. Todo profesional que desee colaborar en la revista deberá atenerse a las siguientes indicaciones:

1. Los trabajos deberán ser originales y versar sobre investigación educativa.
2. El autor o autora deberá enviar un original y tres copias mecanografiadas a la redacción de la Revista.
3. El autor o autora deberá especificar debajo del nombre del artículo su dirección profesional y su E.mail. Si son varios autores/as, se adjuntará el E.mail del primer autor como mínimo.
4. La extensión máxima de los trabajos no deberá exceder las 25 páginas en DIN-A4, a doble espacio y numeradas. Se acompañará un abstract de 100 a 150 palabras en inglés y español.
5. Con la finalidad de simplificar el proceso de confección de la revista y, sobre todo, de reducir al máximo las erratas y costos, se ruega a los autores/as enviar el trabajo, además de las copias en papel, en disquette compatible. Se aceptan los siguientes programas de tratamiento de textos: Word Star, Word Perfect, MSWord y otros.
6. Los trabajos recibidos serán sometidos a informe del Consejo Asesor de la Revista (quien decidirá sobre su publicación o no).
7. Los trabajos recibidos para su publicación en la revista estarán sujetos a un doble proceso de filtro para garantizar la objetividad del juicio. En primer lugar, existirá un proceso de separación de las identificaciones de los autores/as de los trabajos candidatos a ser publicados junto con la asignación de códigos identificativos. Además, existirá un proceso completamente independiente de elección de jueces entre el Consejo Asesor. Ambos procesos serán totalmente independientes y transparentes entre sí, de forma que al asignar un conjunto de jueces del Consejo Asesor se desconoce cualquier información sobre el artículo, salvo el conjunto de descriptores sobre su contenido.
8. Se notificará a sus autores/as los trabajos aceptados para su publicación.
9. Para la redacción de los trabajos se recomienda que sigan las normas de la A.P.A. (American Psychology Association). (Ver adaptación publicada en el n.º 19, 1.er semestre 1992 de R.I.E.).
10. Las fichas resumen se cumplimentarán de acuerdo con el modelo propuesto e impreso por la Revista, a cuya sede podrá solicitar los ejemplares.
11. Los números se cierran los días 1 de mayo y 1 de noviembre de cada año.
12. La R.I.E. es de carácter semestral.
13. Los autores/as deben incluir en sus artículos referencias a otros trabajos publicados en RIE que traten sobre el mismo tema.

REDACCIÓN: REVISTA DE INVESTIGACIÓN EDUCATIVA
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Pedagogía
08035 BARCELONA (Spain)

Sumario

Editorial 3
Eduardo Abalde Paz

ESTUDIO MONOGRÁFICO

La evaluación del profesorado y la gestión de la calidad de la educación. Hacia un modelo comprensivo de la evaluación sistemática de la docencia 7
Juan Mateo Andrés

TRABAJOS DE INVESTIGACIÓN

La opinión de los estudiantes universitarios sobre el método docente de las facultades de ciencias 37
Francisco José Sáez

Diseño de escales para la evaluación de la habilidad motriz de salto para la educación física en Educación Primaria 47
Francisco Javier Castejón Oliva

Efectos de la diferencia de edad y sexo entre tutor y tutorado en una intervención cognitivo-conductual dirigida a mejorar la autoestima, basada en una estructura de tutoría recíproca 69
Miguel Moreno Moreno

Evaluación de un programa de orientación para la inserción laboral en los centros penitenciarios 83
Gemma Filella Guiu

Modelización estocástica de la elección de carrera universitária y de los resultados obtenidos en primer curso de Pedagogía y Magisterio 97
Gómez García J., Hernández Pina F., Buendía Moya F. y Carmona Martínez M.

Autoconcepto, autoestima y rendimiento académico en alumnos de 4º de E.S.O. Implicaciones psicopedagógicas en la orientación y tutoría 119
Miguel Ángel Broc Cavero

La formación inicial del profesor de Educación Secundaria. Situación actual y perspectivas de futuro . 147
Antonio Pantoja Vallejo y Tomás J. Campoy Aranda

Metodología para la enseñanza de las matemáticas a través de la resolución de problemas: un estudio evaluativo 175
Teresa González Ramírez

Análisis y validación de una escala para medir la conducta exploratoria 201
Trinidad Donoso, Pilar Figuera y Mercedes Torrado

TRABAJOS METODOLÓGICOS

Criterios de validez en la investigación cualitativa: de la objetividad a la solidaridad 223
M. Paz Sandín Esteban

FICHAS RESUMEN DE INVESTIGACIÓN 243

8 402124 000683

Volumen 18, número 2, 2000

Revista de Investigación Educativa

Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE)

EVALUACIÓN DE PROGRAMAS EDUCATIVOS

Ramón Pérez Juste (Coordinador)

BEST COPY AVAILABLE

247

ISSN: 0212-4068
Depósito Legal: MU-724-1996

Revista de Investigación Educativa

Volumen 18, número 2, 2000

Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE)
Miembro de la European Educational Research Association (EERA)

Director:
Eduardo Abalde Paz

Directora ejecutiva:
Mercedes Rodríguez Lajo

CONSEJO ASESOR:
Margarita Bartolomé
Nuria Borrell.
Leonor Buendía
Iñaki Dendaluce
José Cajide
Narciso García
Fuensanta Hernández Pina
Jesús Jornet
Joan Mateo Andrés
Mario de Miguel
Arturo de la Orden
Carmen Jiménez Fernández
Antonio Rodríguez Diéguez
Francisco J. Tejedor

CONSEJO DE REDACCIÓN:
Manuel Álvarez
Rafael Bisquerra
Flor Cabrera
Inmaculada Dorio
Julia Victoria Espín
Pilar Figuera
Mª Ángeles Marín
Mercè Noguer
Mª Luisa Rodríguez
Mercedes Rodríguez
Mª Paz Sandín
Trinidad Donoso

SUSCRIPCIÓN E INTERCAMBIO CIENTÍFICO:
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Pedagogía
Paseo del Valle de Hebrón, 171, 2ª planta
08035 BARCELONA (Spain)

DISTRIBUCIÓN:
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Educación
Campus de Espinardo
Universidad de Murcia 30100
Tel. (968) 36 40 67

DISEÑO, MAQUETACIÓN E IMPRESIÓN:
Compobell, S.L. MURCIA

ISSN: 0212-4068

Depósito Legal: MU-724-1996

Web: Rie www.um.es/~dipmide
Aidipe www.uv.es/aidipe

Revista de Investigación Educativa

Volumen 18, número 2, 2000

Presentación 251
Ramón Pérez Juste

PRIMERA PARTE

La evaluación de programas educativos: conceptos básicos, planteamientos generales y problemática 261
Ramón Pérez Juste

La evaluación de programas sociales. Fundamentos y enfoques teóricos 289
Mario de Miguel Díaz

El diseño y los diseños en la evaluación de programas 319
Francisco Javier Tejedor Tejedor

La validez en la evaluación de programas 341
Jesús Jornet Meliá, Jesús M. Suárez y A. Pérez Carbonell

El análisis de datos en la evaluación de programas educativos 357
Luis Lizasoaín Hernández y Luis Joaristi Olariaga

La función optimizante de la evaluación de programas evaluativos 381
Arturo de la Orden Hoz

SEGUNDA PARTE

El muestreo y su problemática en las evaluaciones de programas institucionales 393
José Luis Gaviria Soto

La evaluación y mejora de la enseñanza en la Universidad: otra perspectiva ... 405
Tomás Escudero Escorza

La evaluación de la actividad docente en la Universidad: entre el sueño y la realidad 417
Gregorio Rodríguez Gómez

La evaluación de la enseñanza por los alumnos en el plan nacional de evaluación de la calidad de las universidades. Construcción de un instrumento de evaluación 433
Juan Ruiz Carrascosa

Evaluaciones de programas de formación universitaria en el ámbito europeo y americano. Coherencia con las demandas sociales 447
Honorio Salmerón Pérez

Nuevas tendencias en la evaluación de programas de Educación Multicultural 463
Margarita Bartolomé Pina y Flor Cabrera Rodríguez

249

La evaluación de programas informatizados 481
Víctor Álvarez Rojo, Eduardo García Jiménez, Javier Gil Flores, Soledad Romero Rodríguez y Javier Rodríguez Santero

El programa de orientación «tu futuro profesional» y su evaluación 493
Elvira Repetto y J. Antonio Gil

El orientador como investigador de la acción 509
Rafael Sanz Oro

La evaluación de la formación ocupacional y continua 521
J.M. Jornet, Jesús Suárez Rodríguez y Mª J. Perales

Evaluación de programas para la educación de sobredotados 539
José Cajide Val

Evaluación de programas para alumnos superdotados 553
Carmen Jiménez Fernández

Evaluación de programas para alumnos de alta capacidad: algunos problemas metodológicos 565
Javier Tourón

Evaluación de programas de Educación Emocional 587
Manuel Álvarez, Rafael Bisquerra, Eva Fita, Francesc Martínez y Núria Pérez

Evaluación de programas de atención a los niños con discapacidades 601
María Luisa Dueñas Buey

TERCERA PARTE

Universidad de Alicante 615
Universidad de Barcelona 621
Universidad de Cádiz 627
Universidad Complutense 639
Universidad de Granada 647
Universidad de Jaén 653
Universidad de Málaga 659
Universidad de Murcia 665
Universidad de Navarra 671
Universidad del País Vasco/EHU 681
Universidad de Salamanca 691
Universidad de Santiago de Compostela 701
Universidad de Sevilla 707
Universidad Nacional de Educación a Distancia 713

250

PRESENTACIÓN

El presente número monográfico, dedicado a la evaluación de programas educativos, responde a un encargo de los responsables de AIDIPE, aceptado con gusto e interés por parte de quien escribe.

En mi responsabilidad de editor del número monográfico he deseado compatibilizar dos importantes principios: a) dejar clara mi propia posición, que pretende establecer una íntima relación entre los trabajos pedagógicos y la educación real que se da en los diferentes ámbitos, desde el aula al centro, desde la familia a la sociedad, y b) respetar plenamente las posiciones de todos los compañeros invitados a participar con sus aportaciones. Queda para el lector la tarea de valorar el número como unidad y cada una de las contribuciones en particular.

EDUCACIÓN Y PEDAGOGÍA

Educación y Pedagogía no siempre han ido de la mano, como correspondería a dos realidades hermanadas al ser la primera el objeto de estudio de la segunda. En esa falta de sintonía ha sido frecuente que, quienes cultivamos la Pedagogía como hacer científico, nos hayamos dedicado al estudio de problemas alejados de la realidad que ocupa, y preocupa, a los educadores en las aulas, en el seno familiar o en el ámbito comunitario, o que lo hayamos hecho desde enfoques sociológicos, psicológicos, políticos e incluso económicos, en lugar de abordarlos decididamente desde la perspectiva que nos es propia.

Si lo segundo puede entenderse en la medida en que la realidad educativa es compleja, poliédrica, multidimensional, lo primero encuentra en ocasiones una explicación en el propio clima universitario: hay que estar «a la última», bien sea estudiando el problema de moda, bien acudiendo a la última teoría formulada en un congreso o debate científico, bien utilizando la metodología más avanzada, aunque no tengamos la convicción de que con ello abordamos problemas reales o les damos soluciones adecuadas. En esas condiciones, es frecuente el divorcio entre Educación y Pedagogía, constatado en la mínima influencia de la investigación pedagógica sobre la acción educativa de los profesores en aulas y centros o de cualquier educador en su ámbito de actuación[1].

---

1 En modo alguno pretendo mantener una especie de servidumbre de la Pedagogía a la Educación. La Pedagogía debe, sí, encontrar soluciones rigurosas, científicas si cabe, a los problemas reales de la Educación, pero también debe ir por delante de ella, atisbando problemas, adelantándose a las situaciones, promoviendo innovaciones...

La evaluación de programas o, si se quiere, la investigación evaluativa, es una modalidad de actuación, pedagógica si se enfoca desde tal perspectiva, que puede, y debe, hallarse alejada del primero de los dos defectos ya que, por su propia naturaleza, se encuentra «pegada a tierra» al centrar su objeto de reflexión, de análisis e investigación en los programas educativos en los que bien sean los profesores y educadores, bien los responsables de los sistemas educativos, concretan sus actuaciones sistemáticas al servicio de las metas educativas.

LA EVALUACIÓN DE PROGRAMAS Y LA MEJORA DE LA ACCIÓN EDUCATIVA

Este hecho, por sí sólo, bastaría para reconocer a la modalidad de investigación estudiada en el presente número de RIE una potencial gran importancia para mejorar la realidad educativa. Es por ello que, recientemente, con motivo de un encuentro del área en Valencia[2], afirmaba lo siguiente:

> *La evaluación de programas educativos tal vez sea, o deba serlo, una de las modalidades de investigación pedagógica más utilizadas y, desde luego, con mayores posibilidades y aplicaciones.*
> *No debe extrañarnos que así sea si tenemos en cuenta que la educación, por su propia naturaleza, es una actuación organizada y sistemática al servicio de la mejora o perfeccionamiento del ser humano, concepción totalmente compatible con la de programa, que no es sino un plan sistemático de actuación al servicio de metas pedagógicamente valiosas. Podríamos afirmar que los educadores encarnan sus planteamientos educativos en programas de mayor o menor amplitud, complejidad y duración. Otra cosa será si los programas de los educadores responden a las características técnicas que les son exigibles.*

Pues bien, la evaluación de programas puede realizar grandes aportaciones a la educación, comenzando por esta última: analizar y valorar si «*los programas de los educadores responden a las características técnicas que les son exigibles*». Pero, también, muchas otras, algunas de enorme entidad y valor pedagógico, resaltando las siguientes:

- Si los propios programas son *educativos*.
- Si se da coherencia entre los programas como documentos y su efectiva implementación y desarrollo.
- Si la propia evaluación, **en** y **del** programa, resulta armoniosa y coherente con el programa y con su implantación.
- Si el programa resulta eficaz, efectivo y eficiente.
- Si aparecen efectos no planeados y no deseables.
- Si la evaluación del programa sirve para algo más que la pura constatación de su eficacia, contribuyendo a su continua y progresiva mejora en sucesivas aplicaciones.
- Si ...

---

2 *I Jornadas de Medición y Evaluación Educativas: Estándares e indicadores para analizar la realidad educativa*, celebradas en la Universidad de Valencia los días 8 a 10 de marzo de 2000.

Y es que, realmente, las aportaciones de una adecuada evaluación de programas son muchas, variadas y de gran valor pedagógico, siendo las circunstancias y las necesidades de cada caso las que ordenarán la evaluación en la dirección adecuada. De hecho, los movimientos de calidad, con sus diversos modelos más o menos elaborados e implantados[3], reconocen en la evaluación la principal estrategia o herramienta al servicio de uno de sus principios fundamentales: el de la *mejora continua*. En la misma dirección se manifiesta el Consejo de la Unión Europea[4], cuando recomienda a los Estados «*Apoyar y, en su caso, crear sistemas transparentes de evaluación de la calidad*», al servicio de diferentes objetivos, siendo el segundo de ellos el de

> «*estimular y ayudar a los centros de enseñanza superior para que utilicen medidas adecuadas, en particular en lo que se refiere a evaluación, con el fin de mejorar la calidad de la enseñanza y del aprendizaje, así como de la formación en la investigación, otro ámbito importante de su misión*».

Parece, por tanto, necesario y conveniente que el Área de *Métodos de Investigación y Diagnóstico en Educación* dedique una parte importante de sus esfuerzos a tema de tanta relevancia y trascendencia.

LOS DOS POLOS DE REFERENCIA

La evaluación de programas educativos es, pues, una actividad pedagógica con gran potencialidad de mejora de la acción educativa, pero puede tomar manifestaciones muy diferentes según nos refiramos a ella como una actividad ordinaria de los profesores, a título individual o como equipos docentes, o lo hagamos desde el enfoque de la denominada *investigación evaluativa*.

Si en el primero de los casos no es preciso un gran aparato metodológico, siendo suficiente con una adecuada actitud de autocrítica, una sencilla herramienta para la autoevaluación —individual o en pequeño grupo— y un fuerte compromiso con la mejora, en el segundo la metodología se convierte en un elemento básico.

Pero me interesa dejar clara constancia de que la mejora de la educación en el aula, en el centro, en la familia o en los diferentes ámbitos sociales está mucho más ligada a la efectiva autoevaluación llevada a cabo por los educadores sobre sus respectivos programas que a las evaluaciones rigurosas, a veces sofisticadas, que puedan llevar a cabo los expertos sobre los programas diseñados por las administraciones educativas o aquellas entidades con responsabilidades de diferente índole ligadas a la educación de las jóvenes generaciones[5].

---

3 Destaquemos entre ellos las normas ISO, Calidad Total, Premios Deming o Baldrige, o el modelo europeo de gestión de la *European Foundation for Qualiy Management* (EFQM).

4 En su recomendación de 24 de septiembre de 1998 sobre la cooperación europea para la garantía de la calidad de la enseñanza superior. (98/56 1/ CE).

5 Entre otras razones porque tales evaluaciones no siempre son utilizadas al servicio de la mejora, quedando en ocasiones en actuaciones que acaban en sí mismas, como un requisito socialmente bien visto, pero sin incidencia alguna para futuras aplicaciones de los programas y proyectos.

EVALUACIÓN ORDINARIA DE PROGRAMAS Y PROPUESTAS EVALUATIVAS FACTIBLES

No debe extrañar, en este contexto, que la tarea del Área en el caso de las evaluaciones ordinarias a realizar por parte del profesorado —autoevaluaciones— debiera centrase en la elaboración de propuestas metodológicas sencillas, razonables, factibles en definitiva, que, sin distraer a los profesores de su función: enseñar, formar, educar, les ayuden a hacerlo de modo satisfactorio, eficiente y eficaz, a la vez que les resulten de utilidad para su mejora como profesionales.

La tarea no es fácil, pero debería ser abordada desde el Área; consistiría en la detección y delimitación de los componentes básicos de la evaluación de los programas que, dejando de lado otras manifestaciones secundarias, aportaran el máximo de información relevante con el mínimo de esfuerzo y tiempo por parte de los equipos de profesores.

La elaboración de una tal pauta sencilla, factible y útil, además de la contribución a la creación en los centros educativos y, sobre todo, en los de formación del profesorado, de una *cultura evaluativa*, integrada por una adecuada capacitación y una actitud favorable a la evaluación integral, serían las dos responsabilidades prioritarias del Área.

Conviene no olvidar, por otra parte, que un planteamiento sistemático de la evaluación de programas ordinarios puede dar lugar a un conocimiento técnico de gran relieve y utilidad tanto para la mejora de la actividad educativa como de la metodología de la evaluación de los programas educativos.

INVESTIGACIÓN EVALUATIVA DE PROGRAMAS

La investigación evaluativa ha de responder a los esquemas de rigor de toda investigación, teniendo en cuenta, eso sí, las peculiaridades que definen a la evaluación en relación con la investigación y que fueron expuestas con claridad por Popham[6] al analizar las diferencias en sus objetivos, nivel de generalización y énfasis valorativo.

En relación con el rigor metodológico, me interesa resaltar un hecho de especial relevancia: la *complejidad* y la *multidimensionalidad* de los programas educativos, que debe encontrar su adecuada respuesta metodológica.

La respuesta la planteo en términos del *principio de complementariedad metodológica*, concepto debido al profesor V. García Hoz[7], que va más allá del *multiplismo* formulado por Cook en 1985[8]. El principio implica la tendencia a acudir, y armonizar, las aporta-

---

6 Popham, W.J. (1980). *Problemas y técnicas de la evaluación educativa*. Madrid: Anaya, 20-25. Puede consultarse la entrada *Investigación evaluativa* en A. de la Orden Hoz (1985. *Investigación educativa. Diccionario de Ciencias de la Educación*. Madrid: Anaya.

7 García Hoz, V. (1991). El principio de complementariedad en la investigación pedagógica y en la educación personalizada. *Anales de la Real Academia de Ciencias Morales y Políticas*, 68, 299-317.

8 Cook, T.D. (1985). Postpositivist Critical Multiplisms. En L. Shotland y M.M. Marck (Edits). *Social Science and Social Policy*. Beverly Hills: Sage.

ciones de los diferentes enfoques metodológicos, aplicados a las distintas evaluaciones integrantes de una evaluación integral, las diversas fuentes productoras de información y, por consiguiente, las distintas técnicas e instrumentos para su recogida sistemática y rigurosa, atendiendo a las exigencias técnicas de fiabilidad, validez y valía de la información.

En este punto, resultan de extraordinario valor los estándares de calidad del Joint Committee, publicados en 1994, si bien parece necesario plantear también estándares sobre los propios programas. En tal sentido se sugiere la formulación de estándares sobre la calidad intrínseca, técnica y metodológica de los mismos.

EL NÚMERO MONOGRÁFICO

El número monográfico sobre *Evaluación de Programas Educativos*, se estructura en tres partes claramente diferentes pero íntimamente relacionadas y complementarias entre si:

a) ***Planteamientos metodológicos generales***. Reflexiones y propuestas metodológicas válidas «en general», aunque en cada evaluación concreta tales propuestas puedan presentar matices y diferencias.
b) ***Planteamientos metodológicos especialmente adecuados a la evaluación de cierto tipo de programas***, bien por su mayor o menor extensión, amplitud y complejidad bien por la propia naturaleza o temática de los programas (bien dotados, necesidades educativas especiales, multiculturalidad...).
c) ***Planteamientos que vienen siendo utilizados en la realización de evaluaciones concretas en los Departamentos del Área de MIDE***. No se trata de informes de investigación sino de trabajos de síntesis, realizados por una persona encargada de ello en cada Departamento. Con esta tercera parte se pretende alcanzar algo así como un ***mapa metodológico*** del Área en el marco de la Evaluación de Programas.

Si bien se considera que las aportaciones al apartado a) tratan los problemas fundamentales de su ámbito, los correspondientes al apartado b) no son sino unos ejemplos concretos de los muchos que podrían haberse seleccionado. Al apartado a) corresponden las aportaciones de la primera parte del número monográfico; al b) las de la segunda parte; y al c) las de la tercera parte.

Con estas premisas, la estructura del número y su contenido es la siguiente[9]:

PRIMERA PARTE

1. **La evaluación de programas educativos: conceptos básicos, planteamientos generales y problemática.** Por Ramón Pérez Juste.

9 Es preciso indicar de que algunos autores han modificado el título inicialmente concertado.

2. **La evaluación de programas sociales. Fundamentos y enfoques teóricos**[10]. Por Mario de Miguel Díaz.
3. **El diseño y los diseños en la evaluación de programas.** Por Francisco Javier Tejedor Tejedor.
4. **La validez en la evaluación de programas.** Por Jesús Jornet Meliá, J.M. Suárez y A. Pérez Carbonell.
5. **El análisis de datos en la evaluación de programas educativos**. Por Luis Lizasoaín Hernández y Luis Joaristi Olariaga.
6. **La función optimizante de la evaluación de programas evaluativos.** Por Arturo de la Orden Hoz.

SEGUNDA PARTE

7. **El muestreo y su problemática en las evaluaciones de programas institucionales.** Por José Luis Gaviria Soto.
8. **La evaluación de la enseñanza en la Universidad.**
   8.1. La evaluación y mejora de la enseñanza en la universidad: otra perspectiva. Por Tomás Escudero Escorza.
   8.2. La evaluación de la actividad docente en la universidad: entre el sueño y la realidad. Por Gregorio Rodríguez Gómez.
   8.3. La evaluación de la enseñanza por los alumnos en el Plan Nacional de Evaluación de la Calidad de las Universidades. Construcción de un instrumento de evaluación. Por Juan Ruiz Carrascosa.
9. **Evaluaciones de programas de formación universitaria en el ámbito europeo y americano. Coherencia con las demandas sociales.** Por Honorio Salmerón Pérez.
10. **Nuevas tendencias en la evaluación de programas de Educación Multicultural.** Por Margarita Bartolomé Pina y Flor Cabrera Rodríguez.
11. **Los programas de Orientación y su evaluación.**
    11.1. La evaluación de Programas Informatizados. Por Víctor Álvarez Rojo, Eduardo García Jiménez, Javier Gil Flores y Soledad Romero Rodríguez.
    11.2. El Programa de Orientación «Tu futuro profesional» y su evaluación. Por Elvira Repetto Talavera y Juan Antonio Gil.
    11.3. El orientador como investigador de la acción. Por Rafael Sanz Oro.
12. **La evaluación de la formación ocupacional y continua**, por Jesús Suárez Rodríguez, J.M. Jornet y Mª J. Perales.
13. **Evaluación de programas para la educación de bien dotados.**
    13.1. Evaluación de Programas para la Educación de Sobredotados por José Cajide Val.

---

10 El título inicialmente previsto fue el de *Paradigmas, modelos y metodologías. ¿Hacia la complementariedad metodológica?*

13.2. Evaluación de programas para alumnos superdotados. Por Carmen Jiménez Fernández. Con la colaboración de Leonor Prieto Navarro y de Marta Sandoval Mena.

13.3. Evaluación de programas para alumnos de alta capacidad: Algunos problemas metodológicos. Javier Tourón.

14. **Evaluación de programas de Educación Emocional.** Por Manuel Álvarez, Rafael Bisquerra, Eva Fita, Francesc Martínez y Nuria Pérez.
15. **Evaluación de programas de atención a los niños con discapacidades.** Por María Luisa Dueñas Buey.

TERCERA PARTE

Metodologías habituales en las evaluaciones de programas de los Departamentos del Área MIDE.

Para la tercera parte se invitó a profesores de diferentes universidades para que realizaran una síntesis de las evaluaciones realizadas en los diferentes departamentos del Área de MIDE.

En el presente número se recogen las aportaciones de las siguientes universidades:

Alicante
Barcelona
Cádiz
Complutense
Granada
Jaén
Málaga
Murcia
Salamanca
Santiago de Compostela
Sevilla
UNED

El esquema básico es el siguiente:

TÍTULO DEL PROGRAMA
RESPONSABLE DEL PROGRAMA: Entidad, Director...
RESPONSABLE DE LA EVALUACIÓN: Director, equipo principal...
OBJETO EVALUADO:
Proyecto educativo, proyecto curricular, programa de una o varias materias (conocimientos, procedimientos, estrategias, valores), programa de enseñar a pensar, programa de orientación vocacional, educación vial, educación compensatoria, adaptación curricular, formación ocupacional, formación a distancia...

OBJETIVOS de la evaluación
FUNCIÓN: formativa, sumativa, control, sociopolítica...
PARADIGMA: cuantitativo, cualitativo, crítico, ecléctico (especificar)...
MODELO: resultados, proceso-producto, CIPP...
METODOLOGÍA DE LA EVALUACIÓN:

- **Diseño**
- **Técnicas de recogida de información**
- **Criterios utilizados para valorar la información**: suficiencia, viabilidad, evaluabilidad, eficacia, eficiencia, efectividad, ...
- **Referencias:** estándares prefijados (criterial), normas autonómicas, nacionales o internacionales, superioridad sobre otros programas, progreso sobre el propio programa...
- **Técnicas de registro y análisis de la información**
- **Tipos de decisiones**

INFORME: destinatarios, estructura, amplitud, ...

En algunos de los temas de la segunda parte se invitó a participar a varios profesores que vienen trabajando y, en ocasiones, teniendo responsabilidades diversas, en el ámbito correspondiente. Al tratarse se temas amplios cabía la posibilidad de que la forma en que se abordara el tema fuera diversa, razón por la cual se consideró positivo dar la oportunidad de que tal diversidad pudiera aflorar. En consecuencia, los diferentes invitados han abordado su aportación sin necesidad de establecer contactos destinados a alcanzar la coherencia en los planteamientos; antes al contrario, lo que se pretendió es que quedara patente la diversidad de los mismos en caso de que realmente existiera.

En relación con la tercera parte es preciso indicar que, a pesar del guión de partida, y como se podrá apreciar, la forma en que los diferentes invitados han elaborado su aportación resulta heterogénea, lo que no tiene necesariamente que ser considerado negativo, si bien al editor le hubiera agradado una mayor uniformidad para permitir al lector una mayor facilidad a la hora de tener ante sí ese *mapa metodológico* al que aludíamos más arriba.

Para finalizar, únicamente queda dar las gracias: a los responsables de AIDIPE por el encargo y a todos los invitados que han aceptado por su disponibilidad y su trabajo. Al editor le hubiera gustado contar con más colaboraciones pero, en unos casos por las limitaciones de espacio —el número es ya demasiado amplio— y en otras por no recibir respuesta a la invitación, no ha sido posible. Ahora los lectores tienen la palabra.

Madrid, octubre de 2000

Ramón Pérez Juste

PRIMERA PARTE

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 261-287

LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS: CONCEPTOS BÁSICOS, PLANTEAMIENTOS GENERALES Y PROBLEMÁTICA

Ramón Pérez Juste
UNED

RESUMEN

La necesaria relación entre la Pedagogía y su objeto de conocimiento, la Educación, es objeto de análisis a la hora de plantear las reflexiones pertinentes sobre la evaluación educativa en general y la evaluación de programas educativos en particular.

La evaluación de programas es analizada en la doble perspectiva de la mejora de la acción educativa de profesores y educadores, por un lado, y de la investigación evaluativa por otro.

El autor hace objeto de análisis tres conceptos básicos: programa, evaluación y evaluación de programas, planteando una propuesta metodológica especialmente pensada para la evaluación de programas educativos.

Por último, se ofrece una reflexión sobre algunos temas especialmente críticos, tales como el papel de los expertos, la factibilidad de la evaluación, los aspectos éticos de la evaluación o la comparabilidad de los resultados.

SUMMARY

The necessary connection between Pedagogy and education, as its knowledge object, is the theme of analysis when we put forward accommodated studies on educational evaluation, in general, and on educational programmes evaluation, in particular.

The evaluation of programmes is analysed, on the one hand, with the double approach of the improvement of educational activity implemented by teachers and educators and, from the other one, of evaluative research.

The author offers as object of analysis three basic concepts: programme, evaluation and programme evaluation, by suggesting a methodological proposal particularly directed towards evaluation of educational programmes.

Finally, a consideration is offered on some themes particularly critical, such as the expert´s role, the evaluation feasibility, the ethical implications of evaluation or the results comparability.

1. LA EDUCACIÓN COMO OBJETO DE CONOCIMIENTO E INVESTIGACIÓN

La educación es una actividad radicalmente humana, sistemática, orientada al perfeccionamiento, a la mejora de las personas, de *cada una* de las personas, por medio de acciones intencionadas de los educadores, generalmente concretadas en *planes* o *programas*. Como en cualquier otro ámbito de la realidad natural o social, el ser humano se ha propuesto su conocimiento, bien sea entendido como una *comprensión profunda* de su naturaleza, planteamientos, procesos y resultados, bien como una *explicación* de los mismos mediante leyes más o menos generales, a ser posible de naturaleza causal por ser éste el medio más adecuado para su control y predicción, bien como instrumento al servicio de la *transformación* liberadora de la humanidad[1].

No es mi intención afrontar, con una mínima profundidad, los ya viejos debates sobre la naturaleza de la educación de cara a valorar la adecuación de los diversos paradigmas para producir conocimiento relevante a través de la reflexión y de la investigación. Sin embargo, no es menos cierto que tal planteamiento es básico y que, al menos de forma implícita, subyace en las decisiones que toman los profesionales de la educación o de la Pedagogía cuando abordan este tipo de tareas.

Dejemos clara, eso sí, la compleja y diversa realidad de la Educación como objeto de conocimiento e investigación, de cara a justificar nuestras propuestas metodológicas en un campo de tanto porvenir —si se aborda adecuadamente— como el de la investigación evaluativa o evaluación de programas.

1.1. La educación como *arte*

En el ya añejo discurso sobre estas cuestiones, una de las propuestas es la de considerar a la educación como un *arte*; y no es descabellada la idea sino todo lo contrario, cuando nuestro foco de atención es la relación privilegiada, por única, profunda, intensa y continuada, entre un educador y un educando o todo un grupo de niños o jóvenes en su etapa de formación.

---

1 Entiéndase esta distinción a los efectos meramente expositivos. Difícilmente se puede mantener la incomunicabilidad de estos tres grandes planteamientos sobre la finalidad del saber.

Sin embargo, el hecho de que esa relación sea irrepetible y, en esa misma medida, se acerque a la existente entre el artista y la creación de su obra[2], no impide en modo alguno la posibilidad de que el *artista / educador* encuentre en el saber pedagógico principios, enunciados, propuestas... con los que abordar con mayores posibilidades de éxito una tarea tan apasionante como difícil. Y es que el educador puede aplicar a su situación concreta esos saberes pedagógicos que le indican las regularidades, los grados de dispersión, las tendencias, las secuencias, los ciclos..., al igual que en la pintura, *arte* por excelencia, el pintor aplica de forma creativa los conocimientos ¿científicos? sobre proporciones, perspectiva, contrastes, coloridos ... imprimiéndole después su huella personal.

1.2. La educación como *ciencia*

Precisamente, la educación puede ser objeto de un conocimiento al menos con pretensión de científico, en la medida en que se le aplique un método riguroso al servicio de metas relevantes.

No insistiremos más el manoseado tema de los paradigmas, pero sí dejaremos constancia de que no todos aceptan —en realidad son cada vez menos los que lo hacen— que la única ciencia posible sea la derivada de la aplicación del método hipotético-deductivo-experimental.

Por tanto, en la medida en que el objeto del saber existe, que disponemos de *métodos* rigurosos y que están a la espera múltiples *problemas* y *preguntas* por responder, cabe pensar en una Pedagogía como ciencia, si no con el estatus de las Ciencias Naturales sí con el propio de las de su ámbito, las Ciencias Humanas o Sociales.

En su momento[3] señalábamos como una de las mayores objeciones para aceptar la existencia de una ciencia de la Educación, junto a la relativa al control en los diseños, la posibilidad de establecer leyes generales. Sin embargo, como allí indicábamos, hay autores que defienden tal posibilidad en la medida en que tal concepto no es de significación unívoca. Acudíamos, y lo hacemos de nuevo, a la afirmación de M. Sherwood (*The logic of Explanation in Psychoanalysis*. Academic Press, 1969), que recoge varios tipos de generalizaciones aceptables, reseñando los siguientes:

---

2 Evidentemente, el símil es imperfecto: mientras el artista trabaja con un objeto dócil, estable, que no reacciona a sus intervenciones (o que, cuando lo hace, se comporta de conformidad con las leyes inmutables de la Naturaleza), en educación el objeto / sujeto es capaz de atender o no, de aceptar o no, de poner su esfuerzo o no, y hasta de rechazar las propuestas de sus educadores... El objeto humano es sujeto, con capacidad de tomar decisiones y de comprometerse con ellas con mayor o menor intensidad y esfuerzo. Usamos, por tanto, el símil únicamente en lo que vale: la dimensión creativa de esa relación, donde las reglas, leyes o principios únicamente sirven como referencia y deben ser interpretadas y aplicadas en cada momento en función de las circunstancias.

3 En la Memoria para la oposición de la plaza de profesor agregado en la Universidad de Santiago de Compostela, en 1980.

BEST COPY AVAILABLE

«1 AB se verifica (siempre)
2 En las ocasiones T_1 ... T_n, AB se verifica.
3 En las ocasiones T_1 ... T_n, en las condiciones C_1 ... C_k, AB se verifica.
4 En las ocasiones T_1 ... T_n, en las condiciones C_1...C_k, para los individuos P_1 ... P_l, AB se verifica.
5 AB se verifica en p % de las condiciones (lo que sería la afirmación 1 con la adición de una condición de probabilidad).
6 En las ocasiones T_1 ... T_n, en las condiciones C_1 ... C_k, para los individuos P_1 ... P_l, AB se verifica en el p % de casos.
Siendo AB la relación entre las variables A y B, resulta que la situación 6 es la forma más débil de generalización».

Sin embargo, aun esta «forma débil» de generalización no deja de ser un conocimiento útil para los educadores y, desde luego, adaptado a las características del campo objeto de estudio, donde la diversidad de los seres humanos, y hasta su irrepetibilidad, deben siempre ser tomadas en consideración.

1.3. La educación como *tecnología*

Tampoco es una opción descartable la consideración del estudio de la Educación como una *tecnología*. No olvidemos que según Bunge[4], la *tecnología* nace de aplicar la investigación a mejorar las condiciones de vida del ser humano:

> *«La ciencia como actividad —como investigación— pertenece a la vida social; en cuanto se aplica al mejoramiento de nuestro medio natural y artificial, a la invención y manufactura de bienes materiales y culturales, la ciencia se convierte en tecnología. Sin embargo, la ciencia se nos aparece como la más deslumbrante y asombrosa de las estrellas de la cultura cuando la consideramos como un bien en sí mismo, esto es, como un sistema de ideas establecidas provisionalmente (conocimiento científico) y como una actividad productora de ideas (investigación científica)».*

Parece evidente que en el marco de la educación hay mucho *objeto* especialmente apto y adecuado para un saber *tecnológico*: todo aquello que tiene que ver con la planificación, la programación, el diseño de cursos, de recursos, de materiales, de programas —se destinen al logro de saberes, a la capacitación en destrezas y estrategias o a la creación, desarrollo y mantenimiento de valores—, la organización escolar, los sistemas de evaluación... entran de lleno en es apartado.

4 Bunge, M. (1981). *La ciencia: su método y su filosofía*. Buenos Aires: Ediciones Siglo Veinte, p. 9.

La investigación, y la evaluación, se convierten aquí en instrumentos de enorme importancia para su mejora tanto en la fase de diseño como de implementación, desarrollo y resultados.

1.4. La educación: un campo heterogéneo y multidisciplinar

Las anteriores consideraciones, sobre cuya «corrección» no es fácil decidir dada su indemostrabilidad, si nos llevan, al menos a juicio de quien esto escribe, a una importante conclusión: la educación, una realidad compleja y diversa donde las haya, puede ser considerada desde perspectivas muy diferentes, probablemente todas ellas válidas y necesarias. En cada caso, eso sí, el estudioso deberá saber en qué plano se sitúa y actuar de modo coherente.

Pero hay una segunda e importante cuestión: ¿es la Educación un objeto disciplinar homogéneo? Probablemente muchas de las disputas metodológicas podrían desaparecer o, al menos, atenuarse sensiblemente, si se llegara a una respuesta negativa, esto es, a distinguir entre el estudio científico de la Educación, centrado en la Pedagogía, y las diversas Ciencias de la Educación, encargadas de abordar objetos relacionados con y necesarios para una educación de calidad. Es más, que dentro de la propia Pedagogía como ciencia de la Educación cabe hablar de «objetos» susceptibles de ser abordados multi o pluridisciplinarmente, lo que exigiría que, para alcanzar un acercamiento mínimo y suficiente a tan compleja realidad, se acudiera a unos planteamientos metodológicos acordes con la complejidad y diversidad de su objeto.

Nuestra posición en este punto se concreta en una respuesta negativa a la pregunta y, en consecuencia, en la aceptación de la necesidad de evitar cualquier manifestación reductiva en lo concerniente a la metodología cuando de elaborar conocimiento se trate.

Las dos cuestiones anteriores nos conducen a una misma respuesta en términos de metodología: cualquier propuesta metodológica que se autolimite a una única metodología, además de estar abocada al fracaso por inadecuada a la realidad objeto de estudio, no ofrecerá los frutos apetecibles al dejar fuera de su foco de atención aspectos relevantes de esa compleja realidad. Lo mejor es aplicar el *principio de complementariedad metodológica*, que va incluso más allá de la proposición de Cook[5] conocida como *multiplismo*[6]. El principio de complementariedad reconoce que ante realidades complejas, el mejor acercamiento no reside en la elección de una entre varias opciones, ni siquiera en elegir la mejor, sino en abordarlas desde enfoques metodológicos complementarios[7], siempre que lo permita la formación de los estudiosos, el tiempo y los recursos disponibles.

---

5 Cook, T.D. (1985). Postpositivist Critical Multiplism. En L. Shotland y M.M. Marck (Edits). *Social Science and Social Policy*. Beverly Hills: Sage Pub.

6 En efecto, mientras el *multiplismo* se inscribe en la búsqueda de salidas a las limitaciones del positivismo, el principio de complementariedad metodológica es capaz de dar cabida, en simultáneo o sucesivo, a los métodos cualitativos y cuantitativos, además de a los históricos, filosóficos y comparativos.

7 Podemos encontrar una concreción del *principio de complementariedad metodológica* en el discurso de ingreso del profesor García Hoz en la Real Academia de Ciencias Morales y Políticas. Allí reconoce la riqueza que representa conocer la evolución (historia) de un fenómeno, sus aspectos cuantitativos y cualitativos, su situación en relación con la de otros contextos (sean otros sistemas educativos, otros países...).

1.5. La evaluación de programas y la mejora de la actuación pedagógica

La evaluación de programas es una actividad metodológica que admite, al menos a los efectos expositivos, dos polos separados por un continuo de actuación: la acción reflexiva, ordinaria, de cada profesor o educador sobre su *programa*, entendido como *plan* al servicio del logro de sus metas educativas, y aquella otra llevada a cabo por expertos mediante la aplicación rigurosa de metodologías de diferente naturaleza y alcance, destinada a la evaluación de proyectos y programas de intervención social —en educación, formación, salud, ocio, empleo...— de gran amplitud, complejidad y duración.

En el campo pedagógico, caracterizado por un afán cientifista, los profesores universitarios nos hemos ocupado muy poco de la creación en el profesorado de una auténtica cultura evaluativa sobre sus programas, los que podríamos denominar *programas ordinarios*, a pesar de que es en la aulas, en cada una de ellas, donde se da en mayor grado y con superior intensidad la acción educativa, el hecho educativo.

Crear una auténtica *cultura evaluativa* tendría unos frutos espectaculares; tal cultura supone la existencia de aptitud y de actitud positiva, a lo que se uniría el conocimiento, la información relevante para la mejora que ofrece la evaluación. Al caer en tierra bien abonada por la actitud favorable, una información de tal naturaleza daría sus mejores frutos en el compromiso con las decisiones de mejora de los planes, proyectos, procesos y resultados.

Tal vez sería ésta la mayor aportación a la mejora de la Educación; sin embargo, y en el marco del hacer científico, un adecuado estudio de las actuaciones de los profesores podría ser organizado, sistematizado y tratado desde una perspectiva cercana a la replicación[8], dando lugar al establecimiento de enunciados y conclusiones de cierto nivel de generalidad.

La evaluación de programas, pues, puede servir a esos dos grandes objetivos: la mejora de la calidad de la acción educativa dentro de cada aula y de cada centro, y la creación de teoría, aunque, por prudencia y rigor, demos aquí a tal expresión un nivel atenuado.

1.6. La evaluación de programas como investigación evaluativa

Pero la evaluación de programas es una actividad nacida en otro contexto y orientada hacia otro tipo de objetivos; fundamentalmente nos situamos ante una evaluación externa de programas educativos, realizada por una sola vez, desde el exterior, por expertos, en ocasiones de diversos ámbitos —disciplinares, metodológicos, sociales— respondiendo a lo que conocemos como *investigación evaluativa*. A ella se han dedicado las mejores de nuestras fuerzas ya que se trata de una actividad lucida y recompensada en términos profesionales y, en ocasiones, económicos.

---

8 Vid. al respecto los planteamientos de M. Sidman sobre replicación directa y sistemática en *Táctica de la investigación científica*, Barcelona: Editorial Fontanella, 1978.

Sin embargo, son muchos los que dudan de su valor práctico, de su utilidad, aunque sea forzoso reconocer que los grandes programas institucionales, al igual que las innovaciones de cualquier tipo, suelen llevar unida como una realidad socialmente exigida, una *evaluación de programas*, destinada por lo general a rendir cuentas, a ganar apoyo público (Nevo: 1986) y, al menos teóricamente, a la mejora del programa a través de la toma y aplicación de las adecuadas decisiones de mejora[9].

Lo que se debe pedir a estas evaluaciones es que se realicen con el rigor propio de toda investigación evaluativa, siguiendo hasta donde sea posible y la naturaleza de la situación lo aconseje y lo permita, los estándares propuestos por el Joint Committee[10] y, de modo concreto, en el ámbito de la metodología, aplicando el *principio de complementariedad metodológica* con la mayor riqueza posible.

Vale la pena señalar, por último, que en este ámbito los pedagogos estamos afrontando evaluaciones que no siempre se relacionan en sentido pleno con la educación y que, en consecuencia, muchas de las reflexiones anteriores —y de las que siguen— sólo les serán de aplicación de forma limitada. La ampliación de nuestro campo tanto de intervención como de su evaluación puede responder a causas diversas, no siempre concordantes pero sí perfectamente explicables y muchas defendibles y asumibles. Entre éstas conviene destacar, y apoyar:

a) Aquellas que derivan del reconocimiento profesional de personas que trabajan en el ámbito educativo en general y pedagógico en particular, que les hace ser llamados para que desempeñen, con el rigor y la profesionalidad acreditados, la actividad evaluativa de programas no estrictamente educativos ni pedagógicos.
b) La ampliación del campo de actuación de los pedagogos desde el estrictamente educativo al más amplio de la formación, de gran porvenir e interés en el marco de la *educación permanente*.

Entre las demás, es preciso tomar conciencia de que, en ocasiones, los profesionales de la educación abordamos ámbitos más o menos relacionados con el nuestro como consecuencia, tal vez en simultáneo pero a veces como fenómenos independientes, de, por un lado, nuestra falta de capacidad —o de interés— para desarrollar lo que debe ser nuestro campo de atención preferente y, por otro, del empuje de profesionales de ámbitos cercanos que se hacen cargo de parcelas teóricamente propias de los pedagogos, ya sea por nuestro abandono, ya porque nuestras aportaciones no alcanzan la calidad y los resultados que se esperaban, ya por la indefinición de los

---

9 Forzoso es reconocer que, con frecuencia, y dados los evidentes condicionamientos, estas evaluaciones pueden caer en lo que Stufflebeam y Shinkfield denominan *cuasievaluaciones*, a las que tratan con una enorme dureza al afirmar de quienes las practican que «utilizan la evaluación de una forma inadecuada para lograr sus objetivos», señalando a continuación lo siguiente:

«*Puede que recopilen la información rigurosamente, pero no revelan las verdaderas conclusiones, o lo hacen de un modo selectivo, o incluso las falsifican, engañando tanto a la gente como a sí mismos*» (págs. 67 y 68). Vid. Stufflebeam, D.L. y Shinkfield, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica*, Barcelona: Paidós-MEC.

10 JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994). *The Program Evaluation Standards*, Sage.

respectivos ámbitos de actuación, por no señalar sino algunas de las posibles explicaciones.

Dejemos constancia aquí de que las reflexiones y propuestas que siguen quieren centrarse en ámbitos estrictamente educativos, no por desdén o menosprecio de los demás sino por considerarlos como el campo prioritario de los profesionales de la Pedagogía.

2. CONCEPTOS BÁSICOS

Señalábamos más arriba que el acto educativo tiene mucho de *relación artística* entre el educando y el educador, que trata de aplicar, con la debida prudencia y adecuación[11], los conocimientos científicos, las reglas técnicas y la propia experiencia al aquí y ahora de tal relación.

Sin embargo, el que cada acto educativo tenga o pueda tener ese carácter cuasi artístico no es óbice para que sea conveniente, e incluso necesaria, una acción reflexiva y meditada en la que el educador planifica su acción educativa para el alumno o el grupo, fijando objetivos, seleccionando medios, diseñando las condiciones adecuadas y estableciendo criterios y pautas de valoración[12]. A tal actividad suele dársele el nombre de programación y el de programa a un concepto de muy amplio espectro, pero que, en el campo educativo, no es sino un *plan de acción*, por tanto una actuación *planificada, organizada y sistemática*, al servicio de metas educativas valiosas.

A los efectos de la evaluación de programas conviene clarificar los conceptos fundamentales, destacando entre ellos los de *programa, evaluación* y *evaluación de programas*, a los que nos referiremos a continuación.

2.1. Programa

Como hemos señalado, en el campo pedagógico la palabra *programa* se utiliza para referirse a un *plan sistemático* diseñado por el educador como medio al servicio de las metas educativas. Tanto a los efectos de su elaboración como de su posterior evaluación —dos procesos que deberían guardar armonía y coherencia— deben tomarse en consideración algunas importantes cuestiones:

---

11 No olvidemos que la *educación personalizada* tiene uno de sus principios básicos en la adecuación de la intervención educadora a la diversidad de los educandos, a sus características específicas, a su individualidad —concepto preferido por Allport— tanto en sus aptitudes como en las diferentes dimensiones de su personalidad —desde sus intereses y motivaciones a su modo de ser— y en su capacidad para abordar los diferentes tipos de aprendizaje.

12 La tarea específica del educador consiste en facilitar el éxito de sus alumnos ayudándoles a alcanzar los objetivos planificados. De todos es conocido el planteamiento de Bloom cuando diferencia entre estructura de la disciplina y la estructura del proceso de aprendizaje, «*que debe permitir al estudiante pasar con éxito de una fase de aquél a otra*», concluyendo que «*la estructura de la enseñanza y el aprendizaje deben basarse en consideraciones pedagógicas y no tienen por qué reflejar el punto de vista del especialista sobre su campo de estudio*» (Bloom: 1981, p. 30).

a) Todo programa debe contar con unas *metas y objetivos* que, obviamente, han de ser *educativos*. Esta afirmación parece una obviedad, pero no son pocas las ocasiones en que la evaluación de programas se «olvida» de plantearse esta tan importante como básica cuestión.
b) Metas y objetivos deben estar acomodados a las características de los destinatarios en su contexto de referencia y ser asumidos como propios por los agentes del programa.
c) Tanto a la hora de su implantación como de su evaluación, en este caso en cuanto *variable independiente*, el programa ha de estar claramente especificado y detallado en todos sus elementos fundamentales: destinatarios, agentes, actividades, decisiones, estrategias, procesos, funciones y responsabilidades del personal, tiempos, manifestaciones esperables, niveles de logro considerados *a priori* como satisfactorios ... De no ser así adolecerá de limitaciones en uno de los criterios para su posterior evaluación: la *evaluabilidad*.
d) Ha de incorporar un conjunto de *medios y recursos* que, además de ser *educativos*, deben ser considerados como *suficientes, adecuados y eficaces* para el logro de metas y objetivos.
e) Necesita de un sistema capaz de apreciar tanto si metas y objetivos se logran como de poner de relieve, en caso de que así no sea, o de que lo sea a niveles insuficientes o insatisfactorios, dónde se producen las disfunciones y carencias y, a ser posible, las causas de las mismas (objetivos muy elevados, aplicación inadecuada de los medios, ambiente incoherente, relaciones humanas insatisfactorias, tiempos insuficientes, rigidez en las actuaciones, momentos inadecuados, efectos no planeados...).

Obviamente, tal sistema no es otro que el de la *evaluación*, que debe formar parte de los medios —evaluación *en* el programa— pero que ha de permitir su propia mejora —evaluación *del* programa, como un componente fundamental de su *mejora continua*— además de constituirse, por su propio potencial, en una actividad de mejora profesional[13].

2.2. Evaluación

Aunque es difícil ofrecer una definición que cuente con suficiente consenso y respaldo, a los efectos de este documento el autor la entiende como

> «*...la valoración, a partir de criterios y referencias preespecificados, de la información técnicamente diseñada y sistemáticamente recogida y organizada, sobre cuantos factores relevantes integran los procesos educativos para facilitar la toma de decisiones de mejora*»[14]

---

13 Véase al respecto la aportación de Hopkins (Hopkins, D. (1989): *Evaluation for School Development*. Milton Keynes, Open University Press) cuando distingue entre la evaluación **para** la mejora, la evaluación **de** la mejora y la evaluación **como** mejora.

14 Vid. al respecto Pérez Juste, R. (1995): Evaluación de programas educativos, en A. Medina Rivilla y L.M. Villar Angulo (Edits). *Evaluación de Programas Educativos, Centros y Profesores*. Madrid: Univérsitas, 73-106.

Cuatro son, por tanto, los componentes fundamentales de la evaluación:

a) *Los contenidos a evaluar*: «cuantos factores relevantes integran los procesos educativos». Nuestra posición se concreta en una evaluación integral capaz de armonizar la evaluación de la enseñanza y la del aprendizaje, del alumnado y del profesorado, del programa y de la institución o centro educativo[15].

b) *La información a recoger*: dado su carácter técnico debe ser «técnicamente diseñada y sistemáticamente recogida y organizada». La información es la base de la evaluación, por lo que ésta nunca será mejor que aquélla. En consecuencia, se deberá acudir a cuantas fuentes, técnicas e instrumentos sea preciso para captar de modo suficiente y adecuado cuantos objetos evaluativos se haya decidido evaluar, evitando toda tentación de sesgo o de insuficiencia.
Este punto debe cuidarse con esmero si no se quiere caer en un riesgo especialmente peligroso: el de que la información recogida, por sí misma o por las técnicas e instrumentos utilizados, pueda llegar a desvirtuar los objetivos educativos, que, sin duda, deben ser los elementos rectores tanto del proceso educativo como del propiamente evaluador.
Por último, la información deberá ser organizada de forma que sirva eficazmente a los propósitos de la evaluación, sin llegar a convertirse, como a veces ocurre, en una fase cuasi independiente de la misma, bien por el exceso en su tratamiento bien por su planteamiento al margen de los objetivos a los que sirve.

c) *La valoración de la información*, donde distinguimos los *criterios*, aplicados a las diversas unidades de información, y las *referencias*, utilizadas para emitir juicios globales de evaluación, en nuestro caso sobre el programa. *Criterios* pueden ser los de *exactitud, precisión, claridad, profundidad, variedad...* En cuanto a las *referencias*, las tres fundamentales son las conocidas como *normativa* (baremo), *criterial* y *personalizada o idiosincrásica*.
Vale la pena insistir de nuevo: los *criterios* deben derivar, de modo coherente y armónico, de los *objetivos* educativos a los que sirve el programa; los riesgos derivados de no cumplirse tal exigencia no son otros que los ya apuntados en el apartado anterior, esto es, la desvirtuación de sus objetivos. Algo similar debe afirmarse de las referencias, que deben ser seleccionadas en función de los objetivos y de las finalidades a las que sirve la evaluación[16].

---

15 No estaría de más poder afirmar lo mismo de la evaluación del *sistema educativo*, pero es preciso reconocer que aquí están implicadas dimensiones no estrictamente pedagógicas o educativas, como son las de tipo político, social o económico, por no reseñar sino las más importantes.

16 No parece difícil mantener la superioridad de una evaluación de referencia *personalizada* cuando de lo que se trata es de apreciar el *progreso* educativo de los educandos en pos de su formación personal; o de la *criterial* cuando, además de lo anterior, entran en juego las expectativas y demandas sociales junto a las exigencias de la disciplina objeto de enseñanza y aprendizaje, algo especialmente importante en niveles educativos no obligatorios; o la *normativa*, cuando sea preciso situar los resultados de un alumno en relación con los de compañeros del propio o de otros centros educativos, o los de un centro —incluso de un sistema educativo— en relación con los de otros, sea para situarse en un contexto en que es preciso competir —becas, selección de personal...— sea para orientar decisiones de diversa índole, como ocurre en el caso de los indicadores y estándares nacionales o internacionales...

d) *La finalidad*: «facilitar la toma de decisiones de mejora». Caben otras finalidades o funciones[17], pero, desde una perspectiva estrictamente educativa, la más coherente es la *formativa* o para la mejora, sea del alumno y su aprendizaje, sea del profesor y su actividad educativa, sea del centro educativo como unidad.

Debemos hacer notar que el carácter de herramienta al servicio de los objetivos educativos que atribuimos a la evaluación exige de ésta que sea sea continua y formativa, razón por la que, en un primer ciclo o bucle, deberá sobrepasar la clásica etapa final para alcanzar a los procesos e incluso el diseño y la calidad misma del programa; pero, en realidad, nuestra concepción nos conduce a su institucionalización, con lo que, en ciclos o bucles evaluativos posteriores, se da un enlace y entrelazamiento entre la evaluación final de un ciclo y la inicial del siguiente, lo que permite extraer el máximo potencial de mejora a esta función educativa.

Por último, pero tal vez como lo más importante en tanto que representa una cuestión de principio, deseo resaltar tres características, derivadas de todo lo anterior a la par que su fundamento y soporte, que dan a la evaluación un especial alcance: su carácter *integral, integrado* e *integrador*.

La dimensión o nota de *integral* nos hace ver la necesidad de que la evaluación, además de abarcar la realidad educativa toda, lo haga de modo armónico, al servicio de una gran meta común: la mejora de la persona, bien por su contribución directa a tal objetivo bien por la incidencia que la mejora de los centros, de los programas o del propio profesorado tengan sobre aquella; la nota de *integrada* es una consecuencia de la anterior: la evaluación es un elemento más, junto al resto y, por ello, en necesaria armonía con todos ellos, al servicio de esa finalidad, no debiendo quedar relegada a una actividad añadida, extrínseca y meramente yuxtapuesta; por último, el carácter *integrador* da a la evaluación una dimensión activa y dinámica, y no meramente pasiva, en esa armonización de los medios al servicio de la finalidad de la educación, por tanto en clara contraposición a planteamientos que puedan subvertir el orden, haciendo del instrumento —la evaluación— la finalidad a servir[18], sea por una reordenación de los medios sea por el tiempo —excesivo— que requiera[19].

Obviamente, caben, y así son la mayoría, evaluaciones parciales —sólo los aprendizajes, sólo los programas, sólo los centros, sólo los materiales de aprendizaje...— pero, en tanto que sean evaluaciones pedagógicas, deberían responder a un sistema integralmente concebido y organizado.

---

17 Nevo habla de la *sumativa*, la *sociopolítica* y la *administrativa*.

18 Piénsese, por ejemplo, lo que puede ocurrir cuando de los resultados de una evaluación no centrada en los objetivos educativos puedan seguirse consecuencias no deseables para los educandos, para los profesores o para los propios centros educativos (selectividad, promoción del profesorado, evaluación de centros y sus consecuencias de carácter económico —financiación— o social —certificación y acreditación de la calidad)—.

19 La evaluación no puede consumir un tiempo que deba dedicarse a la acción educativa.

BEST COPY AVAILABLE

2.3. Evaluación de programas

En este marco, entendemos la evaluación de programas como

> *«un proceso sistemático, diseñado intencional y técnicamente, de recogida de información rigurosa —valiosa, válida y fiable— orientado a valorar la calidad y los logros de un programa, como base para la posterior toma de decisiones de mejora, tanto del programa como del personal implicado y, de modo indirecto, del cuerpo social en que se encuentra inmerso»*[20]

Al igual que en el caso anterior, debemos clarificar los conceptos incluidos en la definición, en particular aquellos aspectos que representen alguna peculiaridad respecto de lo ya dicho en el punto 2.2., dedicado a la evaluación en general.

En primer lugar, el *contenido* u objeto de evaluación: sabemos que nos estamos refiriendo a *programas*, pero de todos es conocida la existencia de *modelos* de muy diversa naturaleza en función de su objeto —algunos centrados sólo en los resultados para la constatación del grado de *eficacia*—, de su alcance —evaluaciones por una sola vez— o de sus funciones, fundamentalmente sumativas. Nuestra posición es la de abarcar el programa tanto en su *calidad* —la de sus propios objetivos, de su diseño, de sus medios o del propio sistema de constatación de sus resultados— como en sus *logros* o resultados.

La *finalidad* se centra decididamente en la *mejora*, en perfecta sintonía con la esencia de los actos educativos. Precisamente por esa orientación a la mejora es necesario hacer objeto de evaluación, también, los propios procesos educativos al igual que el proceso de implantación e implementación del programa. Los modelos clásicos centrados en los productos únicamente resultan satisfactorios cuando, tras su aplicación, se constata su eficacia. Pero, en caso de ineficacia, o de niveles insatisfactorios, se revelan incapaces de aportar a los educadores la información necesaria para las decisiones de mejora. En resumidas cuentas, planteamos una evaluación de programas en cuatro grandes etapas —a las que nos referiremos más adelante—, las tres primeras imprescindibles en una primera evaluación, y la cuarta para hacer realidad la integración de la evaluación como una realidad asimilada por el profesorado o la organización.

En cuanto a la *información* ya hemos señalado con anterioridad que es la base de la evaluación, que no puede ser mejor que la información que le sirve de base. Su calidad de *rigurosa* hace referencia a una característica propia de todo procedimiento de acceso científico al saber: no sirve cualquier tipo de información sino aquella que ha sido recogida mediante instrumentos o técnicas debidamente construidos y adecuadamente contrastados en cuanto a sus características básicas de fiabilidad y validez. En este caso, añadimos la nota de *valiosa* para dar a entender la necesidad de que tal información, además, sea *útil* para facilitar la *toma de decisiones de mejora*. Con frecuencia, la naturaleza de la información recogida es lo suficientemente genérica e inespecífica

---

20 Vid. al respecto Pérez Juste, R. (1995): obra citada, p. 73.

como para que las decisiones o no sean adecuadas o, en realidad, no puedan tener como base la información recogida, lo cual haría de la información un componente superfluo[21].

En cuanto a la valoración de la información, daremos por dicho lo señalado al respecto en el punto anterior, señalando, eso sí, la importancia que cobra, dentro del *contexto* en el que el programa se aplica, el *Proyecto educativo* y el *centro educativo*, cuando de programas ordinarios se trate, que, obviamente, deben ser *coherentes* y encontrarse en *sintonía*, en *armonía*, con ambos. La definición y la delimitación de tales contextos, por otro lado, resultan básicas a la hora de establecer tanto los niveles de generalización o, si se prefiere, de aplicación de los resultados a situaciones, ámbitos o contextos diferentes de aquellos en que se aplicó el programa, como de valorar su propia eficacia[22]. Por idénticas razones, en el caso de programas nuevos, diferentes de los ordinarios, será necesario a tales efectos, definir con claridad el contexto de que se trate, sea el familiar, el amical, el social o el colegial.

3. PLANTEAMIENTOS GENERALES: COMPONENTES BÁSICOS DE UNA PROPUESTA PARA LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS

Dada la entidad del concepto *modelo*, tal vez sea preferible acudir a uno menos fuerte, como puede ser el de *propuesta*. Para elaborar un procedimiento completo de evaluación de programas es preciso tomar en consideración, además de lo anterior, algunos elementos con repercusiones metodológicas de diversa naturaleza y alcance, y que, dado el espacio disponible, nos limitaremos a enunciar:

- Lo **educativo** como referencia constante. Tanto en el diseño como en la implementación, desarrollo y evaluación de programas, se deben tener como referencia: que los objetivos, medios y recursos deben ser educativos —no manipuladores— y adecuados; que el programa coexiste con otros programas en el aula y en el centro, con sus correspondientes posibles interacciones; que el *factor humano*, —tanto sus agentes como sus destinatarios, agentes también— es fundamental: sin su decidida implicación y colaboración no hay programa eficaz; o que el contexto, por lo general orientado por «valores» contrarios o, cuando menos, indiferentes a los que subyacen en el programa implantado y evaluado, juega un papel condicionante de primer orden.

---

21 Un ejemplo: en evaluaciones sobre el sistema *a distancia* ha sido habitual preguntar a los alumnos por su preferencia por textos ordinarios o específicos para esta modalidad de aprendizaje, o solicitar su valoración de los textos utilizados en el curso anterior. Piénsese, sin embargo, si en el primer caso no hubiera sido conveniente matizar las respuestas en función de la asignatura de que se trate, de la carrera o del ciclo o curso estudiados; y en el segundo, parece imprescindible acudir a una valoración de *cada* texto ya que una valoración en términos de Media o Mediana, apenas sirve para otra cosa que para quedarse tranquilo si es muy elevada o preocupado en caso contrario. A esta cuestión nos referimos al hablar de información *valiosa*.

22 En ocasiones, un buen programa puede no funcionar adecuadamente por falta de adecuación al contexto de referencia.

- El **tamaño o amplitud del programa:** desde los reducidos —un fin de semana— a los que duran uno o varios cursos, con diferente amplitud y ambiciones, lo que puede alterar sensiblemente la metodología a utilizar[23].
- La **complejidad del programa:** con objetivos de muy diversa naturaleza[24], con las correspondientes implicaciones para su formulación, actuaciones pedagógicas y evaluación.
- El **estatus del evaluador:** como responsable, simultáneamente, del programa y su evaluación, lo que orienta especialmente hacia enfoques autoevaluativos y continuos, centrados en la mejora, o como sólo evaluador, más propio de evaluaciones externas, de expertos, por una sola vez, de carácter sumativo, que concluyen en un informe entregado —previa discusión o no— a los responsables del programa para la toma de las decisiones que correspondan.
- Las **unidades de análisis:** los programas educativos tienen su destinatario natural en las personas de los educandos. Sin embargo, en ocasiones es necesario, bien sea por motivos intrínsecos bien por interés de los responsables del programa o de su evaluación, tomar en consideración otro tipo de unidades, como el aula, el centro, u conjunto de centros (públicos, privados, de educación infantil, de una comunidad autónoma...).
- La relación entre el **programa y su evaluación**: la evaluación debería ser un componente necesario de cualquier programa entendido en forma integral, al menos en su función formativa, pero, con frecuencia, aparece como algo yuxtapuesto o sobrevenido. El hecho de que la evaluación esté o no integrada en el programa condiciona notablemente la metodología, reduciendo su alcance y limitando sus posibilidades en caso negativo.
- La **información a recoger**: dada la complejidad de los fenómenos educativos, una de las tareas más delicadas consiste en la construcción de lo que se considera dato a los efectos de evaluación del programa.
  Vale la pena *exigir* una plena coherencia entre los objetivos del programa, los de su evaluación, y los datos o la información recogida para decidir sobre él; esto, obviamente, no implica desprecio hacia el otro tipo de datos, pero nunca éstos pueden sustituir a aquéllos, sino complementarlos. Y es que, a pesar de la importancia que concedemos a los datos *duros*, dado que la educación se concreta en resultados académicos y, sobre todo, en determinados comportamientos, no se debe olvidar que las *percepciones* de agentes y destinatarios son importantes en la medida en que pueden condicionar la realidad misma del programa. De ahí el valor de formas de triangulación en el análisis de datos.
  Con todo, no son pocas las ocasiones en que los efectos fundamentales del programa son diferidos, por lo que caben evidencias intermedias, aunque sean reco-

---

23 Esta doble dimensión de *tamaño y duración*, no sólo es una cuestión de cantidad sino que, incluso, da lugar a ciertas exigencias cualitativas a la hora de evaluarlo, con repercusiones en temas como la necesidad o no de *muestreo*, tanto de contenidos como de informantes, de estar abierto a posibles efectos *no planeados* o de cuidar determinadas interacciones.

24 Piénsese, en el caso de nuestro actual sistema educativo, la diferencia que representa para la evaluación el que los objetivos sean de tipo *conceptual, procedimental o actitudinal*, o de todos ellos a la vez.

gidas una vez finalizada la aplicación del programa, que pueden tomarse como predictoras de efectos posteriores esperables. Valdría la pena, no obstante, dejar clara constancia de ello y no abandonar su constatación posterior, con frecuencia olvidada porque ya no se rentabilizaría política o socialmente el esfuerzo.

3.1. Componentes de la propuesta

Abordaremos este aspecto desde dos enfoques diferentes, pero complementarios: los correspondientes al *programa* y a su *evaluación*.

A) En relación con el *programa:*

Los elementos a tomar en consideración son:

- El *programa* en cuanto tal y, por tanto, con todos sus componentes: sus objetivos y metas con sus correspondientes contenidos; sus medios y recursos y su evaluación.
- El *programa* en su proceso de implantación y desarrollo.
- El *programa* como realidad llevada a cabo, se piense o no en una nueva aplicación.
- El *programa* como una realidad dinámica: la evaluación debe dar lugar a *consecuencias*, que pueden ser teóricas, prácticas, metodológicas y, sobre todo, educativas, tanto sobre los contenidos educativos como sobre el propio programa.

B) En relación con la *evaluación del programa:*

Desde la perspectiva de la evaluación del programa, sus componentes son:

- La *función* o *funciones* a las que sirve, en cuanto condicionantes del resto de decisiones.
- La *metodología* a utilizar, concepto amplio en el que caben diversos aspectos. Destacamos entre ellos los referentes a la calidad técnica del propio diseño de evaluación del programa; al *diseño* o *plan* general de la evaluación: preguntas a las que responder, objetivos a lograr, niveles a alcanzar (significación estadística, significación pedagógica); *métodos* a utilizar, donde, dadas las consideraciones anteriores, optamos por el *principio de complementariedad metodológica*, bien en forma simultanea bien de modo sucesivo; *información* a recoger en torno al contexto, al programa, a su implantación, a su aplicación, a sus resultados y a sus consecuencias; *momentos* en que habrá de ser recogida la información; *técnicas* e *instrumentos* que habrán de utilizarse; sistema de *registro* y los procedimientos de *análisis; criterios* para valorar cada una de las dimensiones del programa, y *referencias* que permiten una apreciación global del mismo; tipo de *decisiones* previsibles y los procedimientos para su toma, concreción, seguimiento y evaluación.
- El *informe*: su formato, elaboración, debate, difusión.

3.2. Líneas generales de la propuesta evaluativa

La concreción básica de todo lo anterior puede tomar la forma de la siguiente propuesta, estructurada en torno a los cuatro momentos o etapas arriba citadas, siendo las tres primeras fundamentales e imprescindibles en una primera evaluación de un programa educativo. La cuarta representa una situación deseable, según la cual la evaluación se institucionaliza y se integra en sucesivas aplicaciones del programa.

PRIMER MOMENTO: evaluación del programa en cuanto tal

Se trata de la actividad evaluativa más importante tanto por ser la primera y base de todas las demás como por su contenido —abarca el programa en su globalidad— y sobre todo, por sus grandes aportaciones a la mejora y optimización del programa incluso antes de que sea puesto en marcha (función preventiva de la evaluación).

Finalidad:
Establecer la calidad técnica del programa, su viabilidad práctica y su evaluabilidad.
Poner en marcha el programa en condiciones óptimas.

Función:
Formativa: tomar por anticipado las decisiones de mejora que puedan elevar las potencialidades del programa.
En ocasiones puede ser sumativa, sea sobre el programa sea sobre su evaluación.

Metodología:
Análisis de contenido de documentos, estudios prospectivos, técnica Delphi
Juicio de expertos multidisciplinares: metodológicos, científicos, técnicos, pedagógicos.
Registros, pruebas diversas de evaluación inicial (prerrequisitos)

Información a recoger:
Sobre el programa: su fundamentación, su formulación y su relación con las necesidades, carencias, demandas y expectativas de los destinatarios.

Criterios:
Calidad, pertinencia[25] y *realismo* de objetivos y metas y *adecuación* a los destinatarios y al contexto.
Suficiencia de los apoyos, medios y recursos para conseguir los objetivos, en particular sobre la *formación, implicación y compromiso* de los agentes y demás implicados.
Calidad técnica de los planteamientos de evaluación.

25 La *pertinencia social* es un criterio de calidad recogido en la Conferencia Mundial sobre Educación Superior (París, octubre de 1998) en el documento *La Educación Superior en el Siglo XXI. Visión y Acción* (pág. 25). Allí se puede leer: «*La calidad es inseparable de la pertinencia social*», desarrollándose esta idea como «*una búsqueda de soluciones a las necesidades y problemas de la sociedad y más especialmente a los relacionados con la construcción de una cultura de paz y un desarrollo sostenible*».

Calidad técnica del programa: *coherencia* con las bases teóricas y con las necesidades a las que trata de responder; *congruencia* interna entre sus componentes.
Viabilidad del programa.
Evaluabilidad: relevancia, suficiencia, claridad y accesibilidad de la información disponible sobre el programa y de la necesaria en los diversos momentos o etapas de su aplicación y evaluación. Análisis de las dificultades detectadas y previsibles para evaluar el programa

Decisiones:
Generalmente formativas (de mejora previa).
En casos graves, sumativas (retirada del programa, no realización de la evaluación encargada o asumida).

SEGUNDO MOMENTO: evaluación del proceso de implantación del programa.

Finalidad:
Facilitar la toma a tiempo de decisiones de mejora.
Acumular información para introducir mejoras en futuras ediciones del programa.

Función:
Formativa; en casos extremos, sumativa.

Metodología:
Recogida de información: observación, diálogos, entrevistas, análisis de tareas, pruebas formativas intermedias...
Análisis de la información: sesiones de grupo, debates...

Información a recoger:
Sobre el desarrollo del programa, sobre resultados intermedios, sobre efectos no planeados.

Criterios:
Cumplimiento: desfases, desajustes, flexibilidad; *coherencia institucional; eficacia parcial; satisfacción de los implicados:* agentes, colaboradores, destinatarios .

Decisiones:
Ajustes parciales. En casos extremos, suspensión de la aplicación del programa.

TERCER MOMENTO: evaluación de los resultados de la aplicación del programa.

Finalidad:
Comprobar la eficacia del programa

Función:
Fundamentalmente sumativa. Cuando sea posible, sumativa formativizada.

Información a recoger:
Resultados en relación con los objetivos. Efectos —positivos o negativos— no planeados.

Criterios:
Eficacia: grado de logro de los objetivos propuestos. *Eficiencia*: resultados en relación con los medios disponibles y las circunstancias en que el programa se aplica.
Efectividad: efectos beneficiosos no previstos. *Satisfacción* de destinatarios, agentes y otro personal interesado / afectado. *Impacto* del programa en el contexto en que se aplica.

Referencias:
Situación de partida: el programa frente a si mismo **(progreso).**
Niveles preespecificados (evaluación criterial) (**logro del dominio**).
Niveles de otros programas **(superioridad**).

Decisiones:
Sumativas: mantener o suprimir el programa.
Formativizadas: mejorar el programa para una nueva edición.

CUARTO MOMENTO: institucionalización de la evaluación del programa.

Ciclos sucesivos de evaluación / mejora / evaluación / mejora.

En este momento se logra la plena integración del programa y de su evaluación, lo que viene a redundar en la mejora tanto del programa como de la propia metodología de la evaluación, así como en el perfeccionamiento profesional de los educadores.

4. PROBLEMÁTICA

Como se ha podido constatar, la evaluación es una actividad compleja, en la que se cruzan elementos de muy diversa naturaleza, lo que conduce a soluciones muy diferentes y, en ocasiones, muy alejadas unas de otras.

Con todo, la aludida *complejidad* da lugar a una serie de problemas sin resolver, sobre los que habrá que tomar decisiones en cada momento. Reseñamos algunos de los fundamentales:

4.1. La existencia de expertos

Los modelos o propuestas evaluativas externas tienen como base la consideración de la evaluación como una actividad técnica, necesitada de profesionales expertos que la conduzcan a buen puerto.

Con frecuencia, el diseño mismo de la evaluación lo realizan expertos, y personas expertas son las encargadas de recoger, organizar y elaborar la información como base para la configuración de un juicio técnico y hasta para la propuesta de medidas de mejora.

El problema es grave en ciertos tipos de evaluaciones, como lo pone de relieve la situación de la evaluación de la docencia o de la investigación en el Plan de Evaluación de la Calidad de las Universidades, donde no se cuenta ni de lejos con el número de *expertos* necesarios para llevar a cabo tal labor. Pero las consecuencias de una evaluación externa, propia de expertos, llevada a cabo por personas con una formación apresurada, concentrada en una o dos sesiones de fin de semana, son fáciles de imaginar. De por si, la evaluación levanta recelos en las personas e instituciones evaluadas; si a ello se une que quienes han de llevar a cabo la tarea no son reconocidos por los evaluados como dotados de la competencia y de la experiencia necesarias, lo más probable es que la evaluación pierda una gran parte de su efectividad y hasta pueda llegar a ser la gran ocasión perdida: en efecto, cuando los resultados de una evaluación defraudan y son insatisfactorios, y ello tras un largo proceso destinado a crear las actitudes positivas tanto para aceptar la presencia de personas ajenas como para colaborar activamente en las tareas, resulta muy difícil intentarlo de nuevo.

4.2. La factibilidad

En el campo de la evaluación se corren muchos riesgos, y uno de ellos es el intento de perfeccionismo. Cuando un grupo de expertos es invitado a proponer modelos evaluativos, lo hace más desde la perspectiva de sus colegas que de la correspondiente a las personas e instituciones a las que se aplicará la evaluación.

Este hecho suele conducir a propuestas muy elaboradas, amplias, profundas y complejas, cuya aplicabilidad es muy costosa en tiempo, recursos materiales y dedicación del personal.

Nos encontramos ante un importante reto: el de conjugar una evaluación que merezca la pena por la aportación de información que pueda representar, y el tiempo, los recursos y el personal necesarios para lograrlo. La investigación debería orientarse a identificar los contenidos mínimos necesarios para una evaluación que permita ser valorada como *suficiente*, dejando fuera todos aquellos otros aspectos que podríamos denominar *superfluos*. Se trata, en definitiva, de elaborar propuestas centradas en lo esencial, suficientes y factibles.

Recuérdese que las normas del Joint Committee se estructuran en torno a cuatro grandes criterios, siendo el de *factibilidad o viabilidad* uno de ellos, al lado de los de *utilidad, probidad o eticidad y precisión o exactitud*. Los estándares de *factibilidad* se orientan a lograr una evaluación realista, prudente, diplomática y austera, y se concretan en las tres manifestaciones siguientes: *procedimientos prácticos, viabilidad política* y *coste-efectividad*.

La exigencia de factibilidad resulta especialmente necesaria si se quiere crear una cultura autoevaluativa, la que viene considerando el autor como la que representa mayores posibilidades de mejora de la acción educativa en aulas, centros y demás ámbitos educativos.

4.3. ¿Juicios expertos o juicios democráticos?

En la propuesta que hemos ido configurando nos hemos decantado, en lo fundamental, por: a) una evaluación orientada a la mejora; b) desde enfoques internos; inicialmente global, pero, en ciclos sucesivos, parcial, centrando la atención en los puntos *fuertes* y *débiles* previamente detectados; c) realizada por la propia comunidad educativa, con el asesoramiento de expertos cuando no los haya en la propia comunidad; d) aplicando el principio de complementariedad metodológica; e) con la participación de los miembros de la comunidad.

Pues bien, tanto la *participación* como el enfoque de *complementariedad metodológica* conducen a un problema sin resolver: ¿quién o quiénes deben emitir los *juicios* valorativos «definitivos»? y, por tanto, ¿quién ha de tomar las decisiones de mejora?

Lo ideal sería que tanto unos como otros fueran el resultado de *sesiones de evaluación* en las que, con una adecuada dinámica de grupo, los responsables de la institución alcanzaran el consenso en los aspectos fundamentales o, al menos, la aceptación y asunción como propios de los acuerdos mayoritarios. Pero si no fuera así, resolver las cuestiones en forma de votación no parece ser la respuesta más correcta.

Si se acepta que la *participación* es una cuestión de *grado*[26] y que, en consecuencia, admite diversas manifestaciones, desde la simple recepción y aportación de información hasta la co-gestión de las instituciones, pasando por la delegación de competencias..., parece adecuado pensar que, en casos de falta de acuerdo final, sea en el diagnóstico de los puntos fuertes y débiles, sea en la selección de las medidas de mejora a aplicar, sea en la elaboración de los correspondientes planes tanto de mejora de la institución o del programa como de su misma evaluación, las posiciones últimas corresponden a los responsables de su gobierno. No obstante, nunca se insistirá lo suficiente en la importancia de conseguir, mediante el diálogo y las sesiones de grupo debidamente conducidas, un acuerdo básico tanto en el diagnóstico de los puntos débiles como en las medidas a tomar: el apoyo a los planes de mejora depende en gran medida de este importante hecho.

4.4. Los planteamientos éticos

Uno de los problemas más serios de las evaluaciones institucionales, sobre todo cuando es realizada por parte de expertos comisionados o contratados por los responsables del sistema educativo, de los centros educativos o de los programas educativos, es el relativo a la independencia profesional de los evaluadores. Junto a ello, es preciso referirnos al propio comportamiento ético de tales profesionales.

Recordemos al efecto las palabras de Stufflebeam y Shinkfield sobre las pseudoevaluaciones, llegando a denominar charlatanes y manipuladores a quienes no se comportan en este punto de modo adecuado:

---

26 El lector puede consultar Ciscar, C. y Uría, M.E. (1986). *Organización escolar y acción directiva*. Madrid: Popular; o Muñoz, A. y Román, M. (1989). *Modelos de organización escolar*. Madrid: Cincel.

> *«Puede que recopilen la información rigurosamente, pero no revelan las verdaderas conclusiones, o lo hacen de modo selectivo, o incluso las falsifican, engañando tanto a la gente como a sí mismos»*[27].

El problema resultará de difícil solución mientras los evaluadores se encuentren en situación de desigualdad con respecto a quienes encargan las evaluaciones; la falta de asociaciones profesionales fuertes en el campo de la evaluación deja a los evaluadores en la situación de asalariados en relación con sus patronos. Con frecuencia, el evaluador debe optar entre asumir las condiciones de quienes le contratan o arriesgarse a que su trabajo no tenga la difusión necesaria como para influir en la mejora de las personas y de las organizaciones sobre las que ha trabajado profesionalmente.

Este importante problema ha sido recogido en las Normas del Joint Committee, dedicándole uno de sus cuatro grandes apartados, el de *honradez o probidad*, cuya meta se encuentra en asegurar que la evaluación sea realizada legal y éticamente, y con el debido respeto a todos cuantos intervienen y están interesados en su conclusiones. El apartado de *honradez* está integrado por los ocho estándares de calidad siguientes: *orientación al servicio, convenios o acuerdos formales, respeto a los derechos humanos, respeto a la dignidad e integridad humanas, informes completos y claros, declaración de los resultados, conflicto de intereses y responsabilidad fiscal.*

4.5. La comparabilidad de los datos

Cada vez con mayor frecuencia se plantea el problema de la comparabilidad de los resultados, sobre todo cuando se trata de programas en gran escala o de evaluaciones institucionales amplias que afectan al sistema educativo, a las universidades públicas (o a las privadas), a un conjunto de colegios de una misma institución o titularidad... La comparabilidad puede ser una manifestación más de un fenómeno tan reciente como arrollador: el de la globalización.

En ocasiones, este hecho viene impulsado por la función evaluativa de rendición de cuentas o por el cumplimiento del principio de derecho a la información para facilitar las decisiones de los ciudadanos.

El problema de la comparabilidad es de elevada complejidad desde una perspectiva puramente técnica[28], pero se complica todavía más por algunas de las exigencias de nuestro tiempo en lo relativo al *curriculum* de las instituciones o por los derechos consagrados tanto a la libertad de cátedra de los profesores como a la autonomía institucional de las universidades u otro tipo de instituciones.

---

27 Ibídem, págs. 67 s.

28 De hecho, hasta en el marco de los *indicadores* resulta difícil de lograr. Así, en *Análisis del Panorama educativo. Los indicadores de la OCDE. 1995*, editado en Madrid por Mundi-Prensa, se lee lo siguiente: «*Comparabilidad: mejoras y problemas persistentes:* Como resultado de las mejoras en el instrumento e instrucciones de recogida de datos y el continuo diálogo entre la Secretaría de la OCDE y los suministradores nacionales de datos, estos son, de algún modo, comparables. (...) Sin embargo, sigue habiendo importantes problemas de comparabilidad...» (Pág. 11).

De un tiempo a esta parte, en efecto, se está tomando conciencia de la conveniencia de que los *curricula* reduzcan su grado de centralización en los diferentes países, dejando autonomía a las diversas comunidades y regiones en que se estructura cada nación y hasta a los propios centros para acomodar y completar las líneas básicas del curriculo nacional. En España, junto al denominado *Diseño Curricular Base*, de aplicación en todo el país, las diferentes Comunidades Autónomas, en un primer nivel, y los centros educativos, en un segundo, cierran el currículo a cursar por los estudiantes, algo que podrá sufrir adaptación en el nivel del aula. ¿Cómo compaginar este hecho con la necesidad de contar con datos que sitúen los resultados de un país en relación con otros de su entorno? ¿Cómo hacer posible la comparación entre los resultados de diferentes colegios de una misma sociedad o empresa?[29].

El caso de las Universidades puede ser todavía más agudo. Por principio constitucional, las universidades gozan de un elevado grado de autonomía en su organización, gestión y desarrollo de las actividades académicas. ¿Cómo comparar resultados de universidades con tamaño y complejidad muy diferentes, planes de estudio con notable diversidad, número de alumnos, profesores y personal distintos, *ratios* muy alejadas unas de otras...?

Un problema añadido lo es el relativo a las dificultades que este hecho representa para la movilidad estudiantil, tanto para las convalidaciones dentro del propio sistema cuanto para el reconocimiento de Títulos entre diferentes países. Dentro del propio país puede darse el caso de que haya dificultades para poder aplicar una misma prueba para el acceso a la Universidad[30]; las dificultades se acrecientan para homologar los títulos universitarios de diferentes países, incluso dentro de la propia Comunidad Europea.

Una solución, al menos parcial, siempre que la comparación se oriente a la *mejora*, puede encontrarse en seleccionar como referencia no los estándares nacionales o internacionales sino los datos de la propia institución en una perspectiva evaluativa diacrónica. Estaríamos hablando en tal caso de *superación, avance, mejora, progreso* en lugar de *superioridad y competitividad*. Esta perspectiva podría hacerse compatible con una evaluación criterial: la referencia serían los niveles de eficacia, eficiencia, satisfacción... que la propia institución, tras el análisis de su situación, considerara razonablemente alcanzables. Las comparaciones entre los indicadores de los países quedarían limitadas a aquellos aspectos estructurales, funcionales y organizativos en que los datos brutos pueden ponerse en relación con datos contextuales. Por ejemplo: % de alumnos escolarizados en los diferentes niveles; *ratios* de éxito (superación de

---

29 Sobre esta cuestión reflexionó el profesor De la Orden en la conferencia de clausura del XII Congreso Nacional de Pedagogía y I Iberoamericano, celebrado en Madrid en septiembre del año 2000.

30 En España, la prueba de selectividad aplicada tras el Bachillerato para acceder a la Universidad no puede tener como base unas mismas materias debido a que se dan diferencias entre las que se estudian en los diversos cursos. En efecto, las materias que en unas Comunidades Autónomas se estudian en segundo curso de Bachillerato (sobre el que se estructura la citada prueba) en otras se cursan en primero, dándose el caso de que, en otras, se reparten entre primero y segundo.

los diferentes cursos) por etapas educativas; % del producto interior bruto invertido en educación; % de alumnos del Centro educativo que superan las pruebas de selectividad y cuota de estos que culminan, después, los estudios universitarios en cada una de las carreras...

Sin embargo, no estaría de más un elevado esfuerzo de los responsables políticos por encontrar el equilibrio entre la adecuada respuesta a la diversidad social, institucional y personal y el establecimiento de un currículo básico centrado en la cultura propia de los grandes ámbitos territoriales: el Estado, la Comunidad Europea, Occidente.... Ambos polos deben complementarse y nunca darse la espalda.

BIBLIOGRAFÍA

Abraira, C. (1994). *Efectos de la evaluación formativa en alumnos de Matemáticas de Escuela Universitaria de formación de Profesorado de Educación General Básica.* Tesis doctoral. Madrid: UNED.

Aguado Odina, Mª T. (1993). *Análisis y evaluación de modelos y programas de educación infantil en el marco de la Reforma de la Enseñanza.* Madrid: UNED.

Alvira, F. (1991). *Metodología de la evaluación de programas.* Madrid: CIS.

Aparicio, F. y González Tirados, R.Mª (1994). *La calidad de la enseñanza superior y otros temas universitarios.* Madrid: ICE de la Universidad Politécnica.

Bartolomé, M. (1994). La investigación cooperativa. En V. García Hoz (Dir.). *Problemas y métodos de Investigación en Educación Personalizada* (pp. 376-403). Madrid: Rialp.

Bartolomé Pina, M. (1997). Panorámica general sobre Educación Intercultural en Europa. *Revista de Investigación Educativa.* Vol. 15, 1, 7-39.

Brenes Chacón, O.E. (1998). *Influencia de las metodologías activas en la enseñanza de las Ciencias sobre los resultados académicos y socioafectivos.* Tesis doctoral. Madrid: UNED.

Buendía Eisman, L. (1996). La investigación sobre evaluación educativa. *Revista de Investigación Educativa.* Vol. 14, 2, 5-24.

Bunge, M. (1981). *La ciencia: su método y su filosofía.* Buenos Aires: Ediciones Siglo Veinte.

Burgen, A. (Edit) (1999). *Metas y proyectos de Educación superior.* Madrid: Fundación Universidad-Empresa.

Cabrera, F. y Espín, J.V. (1986). *Medición y evaluación educativa.* Barcelona: PPU.

Castillo Arredondo, S. (1987). Evaluación de la eficacia del contenido del consejo orientador. Tesis doctoral. Madrid: UNED.

CECE-Club Gestión de Calidad (1997). *Guía de autoevaluación. Centros Educativos o formativos no universitarios. Modelo Europeo de calidad total en la gestión.* Madrid: ITE-CECE.

CIDE. (1995): *Calidad de la Educación y Eficacia de la Escuela.* Madrid: Autor.

Colás, P. y Rebollo, Mª A. (1993). *Evaluación de Programas. Guía práctica.* Sevilla: Kronos.

Consejo de Universidades (1991). *La evaluación de las instituciones universitarias.* Madrid: Secretaría General del Consejo de Universidades.

Consejo de Universidades (s.f.). *Plan Nacional de Evaluación de la calidad de las universidades.* Madrid: Secretaría General del Consejo de Universidades.

Cook, T.D. (1985). Postpositivist Critical Multiplism. En L. Shotland y M.M. Marck (Edits). *Social Science and Social Policy*. Beverly Hills: Sage Pub.

Cristóbal Martín, M. (1991). *Análisis y evaluación de las prácticas en las Escuelas Universitarias de Formación del Profesorado de EGB de la comunidad de Madrid*. Tesis doctoral. Madrid: UNED.

Dueñas Buey, Mª L. (1988). Análisis y valoración de un ensayo de integración escolar en Madrid-Capital. Tesis doctoral. Madrid: UNED.

Dueñas Buey, Mª L. (1990). Evaluación de programas de apoyo a la integración escolar de deficientes mentales ligeros. *Revista de Investigación Educativa*. Vol. 8, 15, 403-408.

Dueñas Buey, Mª L. (1993). Evaluación de un programa de apoyo a la integración escolar de deficientes mentales ligeros. En P. Colás y M.A. Rebollo. *Evaluación de Programas*. Sevilla: Kronos, 209-214.

Etxeberría, J., García Jiménez, E., Gil Flores, J. y Rodríguez Gómez, G. (1995). *Análisis de datos y textos*. Madrid: Ra-ma.

Fernández Ballesteros, R. (Coord.) (1996). *Evaluación de Programas*. Madrid: Síntesis.

Gamazo Pérez, T. (1999). *Efectos de la aplicación de algunos instrumentos del Programa de Enriquecimiento Instrumental de R. Feuerstein en los educandos de primero y segundo ciclo de Educación Primaria*. Tesis doctoral. Madrid: UNED.

García Hoz, V. (1991). El principio de complementariedad en la investigación pedagógica y en la Educación personalizada. *Anales de la Real Academia de Ciencias Morales y Políticas*. 68, 299-317.

García Hoz, V. (Dir). (varios años). *Tratado de Educación Personalizada*. 33 volúmenes.

García Jiménez, E. (1994). Investigación etnográfica. En V. García Hoz (Dir.). *Problemas y métodos de Investigación en Educación Personalizada* (pp. 343-375). Madrid: Rialp.

Gaviria, J.L. (1999). El enfoque multinivel en la evaluación de sistemas educativos. *Revista de Investigación Educativa*. 17, 2, 429-443.

Gento Palacios, S. (1994). *La gestión de calidad en Educación*. Madrid: La Muralla.

Gento Palacios, S. (1998). *Implantación de la calidad total en las instituciones educativas*. Madrid: UNED.

Ginés Mora, J., I.V.I.E. y Universidad de Valencia (1998). *Indicadores de gestión y rendimiento para las universidades españolas*. Valladolid.

Hernández Pina, F. (1993). Concepciones en el estudio del aprendizaje de los estudiantes universitarios. *Revista de Investigación Educativa*. 22, 117-1545-24.

House, E.R. (Edit). *New Directions in Educational Evaluation*. Lewes: Falmer Press.

INCE (1998). *Diagnóstico del Sistema Educativo 1997*. Madrid: Ministerio de Educación y Cultura.

INCE (2000). *Sistema Estatal de Indicadores de la Educación*. Madrid: MECD.

INCE (2000). *Evaluación de la enseñanza y el aprendizaje de la lengua inglesa. Educación primaria 1999. Avance de resultados*. Madrid: Ministerio de Educación y Cultura.

INCE (2000). *Evaluación de la educación primaria. Datos básicos 1999*. Madrid: Ministerio de Educación y Cultura.

Jiménez Fernández, C. (2000). Cómo conocer a los más capaces: diagnóstico de la alta capacidad. En C. Jiménez Fernández: *Diagnóstico y educación de los más capaces*. Madrid: UNED.

Joaristi Olariaga, L. y Lizasoaín Hernández, L. (1998). *BMDP. New System para Windows*. Madrid: Paraninfo.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*. Thousand Oaks, CA.: Sage.

Jornet Meliá, J. y Suárez Rodríguez, J. (1994). Evaluación referida a criterio. Construcción de un test criterial de clase. En V. García Hoz (Dir.). *Problemas y métodos de Investigación en Educación Personalizada* (pp. 419-443). Madrid: Rialp.

Hopkins, D. (1989). *Evaluation for School Development*. Milton Keynes: Open University Press.

Kemmis, S. (1989). Seven Principles for Programs Evaluation in Curriculum Development and Innovation. En E. R. House (Ed.). *New Directions in Educational Evaluation* (pp. 117-140). Lewes: Falmer Press.

Luján, J. y Puente, J. (1996). *Evaluación de centros docentes. El Plan EVA*. Madrid: MEC.

Marchesi, A. y Martín, E. (1998). *Calidad de la enseñanza en tiempos de cambio*. Madrid: Alianza Editorial.

Martín, Mª T. y Sarrate, Mª L. (1999). *Evaluación y ámbitos emergentes en animación sociocultural*. Madrid: Sanz y Torres.

Martínez Mediano, C. (1991). *Factores de eficacia de los Centros Asociados de la UNED en el curso de acceso para mayores de veinticinco años*. Tesis doctoral. Madrid: UNED.

Martínez Mediano, C. (1996). *Evaluación de programas educativos*. Madrid: UNED.

Mateo Andrés, J. (2000). La evaluación del profesorado y la gestión de la calidad de la Educación. Hacia un modelo comprensivo de evaluación sistemática de la docencia. *Revista de Investigación Educativa*. Vol. 18. 1, 7-36.

Mateo, J., Escudero, T., Miguel, M. de, Ginés Mora, J. y Rodríguez Espinar, S. (1996). La evaluación del profesorado. Un tema a debate. *Revista de Investigación Educativa*. Vol. 14, 2, 73-94.

Michavilla, F. (Edit). (1998). *Experiencias y consecuencias de la evaluación universitaria*. Madrid: Fundación Universidad - Empresa.

Miguel Díaz, M. de (Coord) (1999). La evaluación de programas: entre el conocimiento y el compromiso. *Revista de Investigación Educativa*, Vol. 17, 2, 345-380.

Miguel Díaz, M. de (1997). La evaluación de los centros educativos. Aproximación a un enfoque sistémico. *Revista de Investigación Educativa*. Vol. 15, 2, 145-178.

Ministerio de Educación y Cultura (1997). *Modelo Europeo de Gestión de Calidad*. Madrid: Centro de Publicaciones. Secretaría General Técnica.

Medina Heredia, Mª T. (1998). *Evaluación de programas educativos. Los materiales didácticos del Centro Nacional de Educación a Distancia*. Tesis doctoral. Madrid: UNED.

Mozas Flores, J.C. (1997). *Diseño, aplicación y evaluación de un programa integrado en el currículo de Educación Primaria, sobre el fomento de actitudes contrarias al tabaco y al alcohol en un marco interactivo con los principios de una vida sana*. Tesis doctoral. Madrid: UNED.

Nevo, D. (1986). The Conceptualization of Educational Evaluation. En E.R. House (Edit): *New Directions in Educational Evaluation*. (pp. 15-30). Lewes: Falmer Press.

Nevo, D. (1990). Role of the Evaluator. En H. J. Walberg y G. D. Haertel (Edits.). *The International Encyclopedia of Educational Evaluation*, (pp. 89-90). Oxford: Pergamon Press.

OCDE. (1991). *Escuelas y calidad de la enseñanza. Informe internacional*. Madrid: Paidós-M.E.C.

OCDE (1995). *Análisis del panorama educativo*. Madrid: OCDE-Mundi-Prensa.

Orden Hoz, A. de la (1993). Influencia de la evaluación del aprendizaje en la eficacia de la enseñanza. *Revista de Investigación Educativa*. 22, 7-42.

Orden Hoz, A. de la (1999). La investigación educativa: naturaleza, funciones y ambigüedad de sus relaciones con la práctica y la política educativas. *Revista de Investigación Educativa*. Vol. 17, 1, 7-32.

Orden Hoz, A. de la (2000). La globalización y los estándares nacionales de rendimiento educativo. Conferencia de clausura del XII Congreso Nacional de Pedagogía y I Iberoamericano. Documento no publicado.

Patton, M.Q. (1980). *Qualitative evaluation methods*. Beverly Hills: Sage.

Pérez Juste, R. y García Ramos, J.M. (1989). *Diagnóstico, evaluación y toma de decisiones*. Madrid: Rialp.

Pérez Juste, R. (Dir) (1989). La Universidad Nacional de Educación a Distancia: luces y sombras de una institución innovadora. *Bordón*. 41, 3. Monográfico.

Pérez Juste, R. (1989). Evaluación del material didáctico. *Revista Iberoamericana de Educación Superior a Distancia*. Vol. I, 3, 93-112.

Pérez Juste, R. (Dir) (1991). *La Universidad Nacional de Educación a Distancia: aproximación a la evaluación de un modelo innovador*. Madrid: CIDE.

Pérez Juste, R. (Dir) (1992). *Estudios sobre el curso de acceso a la UNED para mayores de 25 años*. Madrid: UNED.

Pérez Juste, R. (1993). Evaluación de las adaptaciones curriculares. *Revista de Ciencias de la Educación*. 153, 51-68.

Pérez Juste, R. (1993). Evaluación de Programas de Orientación. *Actas del V Seminario Iberoamericano de Orientación*, págs. 46-65.

Pérez Juste, R., Martínez Mediano, C. y Rodríguez Gómez, G. (1994). *Centros educativos de calidad*. Madrid: ITE de CECE.

Pérez Juste, R. (1994). Investigación y mejora de programas. Programas comunes y programas personales. En V. García Hoz (Dir.). *Problemas y métodos de Investigación en Educación Personalizada* (pp. 510-536). Madrid: Rialp.

Pérez Juste, R., García Llamas, J.L. y Martínez Mediano, C. (Coords.) (1995). *Evaluación de Programas y Centros Educativos*. Madrid: UNED.

Pérez Juste, R., Cristóbal Martín, Mª. y Martínez Aznar, B. (Coords.) (1995). *Evaluación de Profesores y Reformas Educativos*. Madrid: UNED.

Pérez Juste, R. (1997). La evaluación del material impreso como elemento sustancial en la evaluación de programas. Ponencia presentada a las *Jornadas Internacionales sobre el Material Impreso en la Enseñanza a Distancia*. Madrid: IUED de la UNED.

Pérez Juste, R. (1998). La calidad de la Educación universitaria. *Revista Iberoamericana de Educación a Distancia* 1, 1, 13-38.

Pérez Juste, R. (1999). Nuevas perspectivas de gestión, control y evaluación. *Bordón*, 51, 4, 449-478.

Pérez Juste, R. (2000). La calidad de la Educación. En R. Pérez Juste y otros: *Hacia una Educación de Calidad*, (pp. 13-44). Madrid: Narcea.

Pérez Juste, R. (2000). La evaluación como medio para la mejora de la calidad y de la eficacia del aprendizaje, de la educación y de las instituciones. Tercera ponencia de la Sección VI del XII Congreso Nacional de Pedagogía. Madrid.

Repetto Talavera, E. (1994). Últimas aportaciones en la evaluación de programas de orientación metacognitiva de la comprensión lectora. *Revista de Investigación Educativa*. 23, 314 -323.

Rodríguez Espinar, S. (1997). La evaluación institucional universitaria. *Revista de Investigación Educativa*. Vol. 15, 2, 179-216.

Rodríguez Gómez, G. (1995). El centro escolar como objeto de investigación: la evaluación interna de centros. En V. García Hoz (Dir.). *Del fin a los objetivos de la Educación Personalizada* (pp. 308-332). Madrid: Rialp.

Rodríguez Gómez, G., Gil Flores, J. y García Jiménez, E. (1996). *Metodología de la investigación cualitativa*. Málaga: Aljibe.

Rossi, P.H. y Freeman, H.E. (1989). *Evaluación: Un enfoque sistemático para programas sociales*. México: Trillas.

Sanz Moreno, A. (1992). *Diseño y evaluación de un programa para desarrollar las habilidades implicadas en la comprensión y síntesis de textos expositivos*. Tesis doctoral. Madrid: UNED.

Sanz Oro, R. (1990). *Evaluación de programas en orientación educativa*. Madrid: Pirámide.

Scriven, M. (1967). «The Methodology of Evaluation», en *Perspectives on Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation, Chicago, Rand McNally.

Sidman, M. (1978). *Táctica de la investigación científica*. Barcelona: Editorial Fontanella.

Stenhouse, L. (1984). *Investigación y desarrollo del curriculum*. Madrid: Morata.

Stufflebeam, D.L. y Shinkfiel, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós-MEC.

Tejedor, F.J. (1990). Perspectiva metodológica del diagnóstico y evaluación de necesidades en el ámbito educativo. *Revista de Investigación Educativa*. Vol. 8, 16, 15-37.

UNESCO (1998). *La Educación Superior en el Siglo XXI. Visión y Acción*. París: UNESCO.

Varios (1991). Evaluación de Programas. *Bordón*, 43, 4, monográfico.

Varios (1998). La evaluación de la Universidad Española. *Revista de Educación*, 315. Monográfico.

Varios (1996). Evaluación de la calidad de la Educación. *Revista Iberoamericana de Educación*, 10, Monográfico.

Villar Angulo, L.M. y Medina Rivilla, A. (Edits.) (1998). *Evaluación de Programas Educativos, Centros y Profesores*. Madrid: Univérsitas.

Wilson, Jh. D. (1992). *Cómo valorar la calidad de la enseñanza*. Madrid: Paidós-M.E.C.

Woods, P. (1987). *La escuela por dentro*. Madrid: Paidós-M.E.C.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 289-317

LA EVALUACIÓN DE PROGRAMAS SOCIALES: FUNDAMENTOS Y ENFOQUES TEÓRICOS

Mario de Miguel Díaz
Universidad de Oviedo

1. INTRODUCCIÓN

Frecuentemente constatamos que en la literatura relativa a evaluación se utilizan algunos conceptos de forma poco precisa lo que genera indudablemente expresiones confusas. Por ello, antes de abordar el tema que nos ocupa, consideramos procedente efectuar ciertas precisiones terminológicas que permitan clarificar el punto de vista que asumimos al realizar este trabajo, así como los parámetros dentro de los cuales debe ser analizada nuestra aportación ya que, de lo contrario, su lectura puede dar lugar a interpretaciones muy distintas a nuestros planteamientos.

Un ejemplo de lo que acabamos de señalar sucede con los términos investigación evaluativa y evaluación de programas. Aunque inicialmente aluden a un mismo campo de trabajo, en sentido estricto, su significado es distinto. Desde nuestro punto de vista, la investigación evaluativa (*evaluation research*) constituye la aplicación de los principios y procedimientos de la investigación social para comprobar sistemáticamente la eficacia de un programa de intervención social. Cuando hablamos de evaluación de programas (*program evaluation*) aludimos al conjunto de principios, estrategias y procesos que fundamentan la evaluación de toda acción o conjunto de acciones desarrolladas de forma sistemática en un determinado contexto con el fin de tomar las decisiones pertinentes que contribuyan a mejorar las estrategias de intervención social. Si el objetivo de un programa es mejorar las condiciones sociales, el propósito de la evaluación es mejorar los propios programas (Rossi, Freeman y Lipsey, 1999). De lo cual se deduce que esta tarea conlleva «algo más» que la mera utilización de metodologías científicas orientadas a evaluar teorías e impactos.

Para percibir con mayor claridad los matices que distinguen ambos términos basta recordar la definición comúnmente aceptada sobre el concepto de evaluación como

«un proceso sistemático de búsqueda de evidencias para formular juicios de valor que orienten la toma de decisiones». De ella se desprende la existencia de tres elementos claves que delimitan esta actividad: *el proceso metodológico que se lleva a cabo para recoger e interpretar las evidencias relacionadas con el programa, los criterios que vamos a utilizar para emitir los juicios de valor y, finalmente, la utilización que posteriormente se va hacer de los resultados*. Por ello, toda aproximación conceptual al campo de la evaluación de programas debe contemplar necesariamente los tres pilares sobre los que se fundamenta esta disciplina —evidencias, valores y decisiones— ya que, de lo contrario, efectuaríamos un reduccionismo injustificado.

Partiendo de este planteamiento, la distinción entre investigación evaluativa y evaluación de programas se percibe de forma más nítida. *La investigación evaluativa constituye esencialmente una estrategia metodológica orientada a la búsqueda de evidencias respecto a un programa, pero la evaluación de un programa requiere, además, abordar otros problemas relacionados con los criterios a utilizar para emitir los juicios de valor y con los procesos relativos a la toma decisiones*. Se diría, pues, que la primera constituye una de las herramientas metodológicas básicas a utilizar en una de las fases del proceso evaluativo. Ciertamente una fase muy importante —en la medida que nos facilita el soporte a partir del cual vamos a emitir las valoraciones— pero en modo alguno la única.

Hemos considerado oportuno iniciar el presente trabajo matizando esta distinción porque en la bibliografía al respecto es bastante común que a la hora de hablar de paradigmas, modelos o enfoques sobre evaluación de programas se utilicen —más o menos adaptados— aquellos que son propios de la investigación científica social como si ambos términos fueran equivalentes. Desde nuestro punto de vista esta extrapolación adolece de sentido; dado que supone una utilización indiscriminada de conceptos que aluden a dominios diferentes. *No cabe, por tanto, trasladar a la evaluación de programas los supuestos teóricos de los paradigmas clásicos de investigación propios de las ciencias sociales ya que la fundamentación de esta disciplina requiere necesariamente contemplar el mundo de los valores y de las consecuencias*. Por ello, necesitamos abordar este tema desde una óptica más comprensiva.

2. DELIMITACIÓN DEL CONTEXTO DEL TRABAJO

Una de las características dominantes en el campo de la evaluación de programas durante los últimos años ha sido la multitud y heterogeneidad de los conceptos asociados con este término. Al igual que sucede en toda disciplina emergente, constantemente están apareciendo aproximaciones, enfoques, modelos y métodos que acuñan nuevos términos y tipologías lo que hace patente la escasa consolidación epistemológica que presenta actualmente este campo. Presionados por la realidad práctica, los evaluadores se dedican a resolver cuestiones de carácter metodológico y relegan a un segundo plano las cuestiones teóricas relativas a los fundamentos de la evaluación. De ahí que *la escasa atención que se presta a la «teoría» se considere como una de las debilidades mas importantes que tiene esta actividad y que impide su consolidación como disciplina científica* (Bickman 1990, Rebien, 1997).

Ante esta carencia, resulta necesario revisar los supuestos teóricos de esta disciplina con el fin de contribuir a consolidar su estatuto epistemológico y ayudar a los evaluadores a que puedan explicitar los criterios que implícitamente utilizan para seleccionar, organizar y llevar a cabo sus métodos de trabajo. La tarea del evaluador no puede limitarse a una simple elección de metodologías y métodos para obtener evidencias. Necesitamos de manera imprescindible una teoría que justifique y explique los procedimientos y las prácticas que en cada caso se llevan a cabo. *La evaluación de programas es algo más que una metodología aplicada y, al igual que sucede en todas las disciplinas, requiere sus propias bases teóricas que justifiquen y den sentido a lo qué hace un evaluador, cómo lo hace y, sobre todo, para qué.*

Como ya hemos avanzado, en este campo no abundan trabajos de carácter epistemológico. Se habla mucho de métodos y metodologías pero casi nada de epistemología; es decir, sobre la teoría del conocimiento vinculada a la evaluación de programas. Más aún, en muchos textos no se matizan las diferencias entre fundamentos científicos de la evaluación de programas (*foundations*) y enfoques teóricos aplicables a la evaluación de programas (*approaches*). Evidentemente estos términos no significan lo mismo ya que, mientras el primero hace referencia a las bases que fundamentan la evaluación, el segundo alude claramente a las distintas metodologías que podemos utilizar en la práctica para abordar los procesos. Aunque, lógicamente, ambos conceptos están relacionados.

Por ello, aunque explícitamente no se manifieste la justificación teórica, cada opción metodológica ha sido elegida en función de las concepciones ontológicas y epistemológicas que el evaluador tiene sobre los conceptos de «programa» y «evaluación», lo que significa que existe un posicionamiento implícito. Detrás de cada aproximación metodológica siempre hay un enfoque teórico tácito. Ello permite que, ante un problema concreto, sea posible plantear estrategias de intervención distintas y que evaluaciones con enfoques diferentes sobre una misma intervención puedan llegar a resultados contradictorios (Dennis y Boruch, 1989). Lo cual significa que *todas las aproximaciones tienen sus ventajas e inconvenientes y que es necesario que el evaluador precise, en cada caso, el marco conceptual que asume a la hora de justificar y planificar su trabajo ya que las decisiones metodológicas deben estar siempre supeditadas a los planteamientos teóricos* (Shadish, Cook y Leviton, 1991). Necesitamos abordar la reflexión epistemológica sobre el tipo de conocimiento que nos aporta un proceso evaluativo y, en consecuencia, sobre el marco que justifica el trabajo de un evaluador. Necesitamos más y mejor teoría (Cook, 1997).

Situados en este contexto, el abanico de temas que pueden ser objeto de la reflexión es tan amplio que necesariamente procede cierta delimitación con fines didácticos. Desde nuestro punto de vista, siguiendo pautas de otros autores que se ocupan del tema (Bickman 1990, Shadish, Cook y Leviton 1991, Rebien 1997), la fundamentación teórica de la evaluación de un programa exige una toma de postura, al menos, sobre las siguientes cuestiones básicas: *el concepto de programa como estrategia de intervención social, las bases epistemológicas que determinan la construcción de las evidencias, los criterios a utilizar para formular los juicios de valor, la utilidad o usos que van a tener los informes y el enfoque metodológico a seguir para realizar el proceso evaluativo.* A partir de las opciones y

decisiones que el evaluador asuma en torno a estas cuestiones podremos entender y justificar las razones por las que elige un determinado enfoque teórico y metodológico para evaluar cada programa concreto.

3. CONCEPTO DE INTERVENCIÓN SOCIAL

Toda aproximación al estatuto científico de una disciplina conlleva identificar previamente cual es el objeto del conocimiento. Necesitamos, pues, una definición sobre el concepto «programa». Habitualmente se entiende por programa todo «*plan, proyecto o servicio mediante el cual se diseña, organiza y pone en práctica un conjunto de acciones dirigidas a la consecución de una metas*» (De Miguel, 2000b). Este término aglutina, por tanto, cualquier tipo de estrategia de intervención social orientada a promover cambios. De ahí que la primera reflexión que se debe plantear el evaluador sea sobre cómo deben ser conceptualizadas, diseñadas y aplicadas estas estrategias de intervención para que produzcan cambios sociales efectivos. En definitiva, necesitamos una teoría que fundamente la elaboración y aplicación de las estrategias de intervención social y, en consecuencia, que pueda ser utilizada como marco de referencia para tomar decisiones a la hora de planificar los procesos evaluativos (Bickman, 1987).

El término «teoría» aplicado a la evaluación de programas alude a «*un conjunto de principios interrelacionados que explican y predicen la conducta de una persona, grupo u organización sobre la que hemos llevado a cabo una acción intencional y sistemática*» (Chen, 1990). Lógicamente, existen distintas formas de entender las relaciones entre los principios, las acciones y las conductas en el campo social, por lo que es muy importante conocer cuál es el punto de vista que considera el evaluador en cada caso concreto, ya que constituye el determinante inmediato que justifica la perspectiva o enfoque teórico que adopta para planificar y evaluar una intervención. De ahí que una de las primeras cuestiones que el evaluador debe explicitar sea la concepción teórica desde la que aborda el «programa» en tanto que estrategia de intervención social.

3.1. El programa como verificación de teorías

Una de las concepciones clásicas sobre los programas es considerarlos como estrategias que permiten la verificación de teorías que son útiles para resolver problemas sociales. Quienes adoptan esta postura enfatizan el papel de la teoría como fundamento de toda intervención social por lo que el objetivo prioritario de toda evaluación se orienta a probar las relaciones causa-efecto que se establecen entre unos tratamientos y unos resultados. Ello exige que tanto el diseño como la ejecución de los programas se lleve a cabo de acuerdo con los principios de la metodología científica más rigurosa dado que su objetivo es muy similar a la verificación de hipótesis conceptuales.

Quienes adoptan esta postura —como ya proponía el conocido trabajo de Campbell (1969), *Reforms as Experiments*—, consideran que los cambios sociales pueden ser planificados a modo de experimentos. Los programas constituyen estrategias o artefactos a través de las cuales se someten a prueba si determinadas teorías son eficaces para la resolución de problemas sociales prácticos. Al igual que sucede en todo experi-

290

mento, el evaluador formula una teoría explicativa sobre una realidad o problema social sobre la cual construye y aplica una estrategia de intervención siguiendo los cánones de la metodología científica (aleatorización y manipulación) con el fin de poder probar su adecuación y eficacia.

Para quienes defienden este concepto, el objetivo principal de una evaluación es lograr identificar los efectos netos de un programa —es decir, aquellos que se deben exclusivamente a las variables explicativas que justifican la intervención— al margen de los efectos brutos. Otros autores prefieren hablar de resultados e impactos, entendiendo por los primeros todos los efectos de un programa y acotando el segundo término solamente para aquellos sobre los que se tenga indicios razonables de causalidad (Schalock 1995, Boruch 1997, Apodaca 1999). Precisamente porque su finalidad primordial es la identificación de los efectos netos o impactos —a corto, medio o largo plazo— asumir este enfoque implica introducir en la evaluación de un programa toda la problemática de la experimentación científica en el campo social, con toda la carga de dificultades y limitaciones que este hecho conlleva. Las características implícitas en toda intervención social impiden la aplicación de diseños metodológicos rigurosos que nos permitan obtener indicios razonables de validez interna y externa de forma satisfactoria (Cronbach, 1982).

3.2. El programa como tecnología sistemática

La realidad práctica nos hace ver que la resolución de problemas sociales constituye más un problema de contingencias que de teorías causales. Generalmente no existen grandes discusiones sobre las teorías o principios que deben inspirar las políticas públicas. Todos sabemos qué procede hacer cuando se quieren abordar problemas relativos a la pobreza, la marginación, la discriminación social, tercera edad, etc... Las diferencias no están tanto en el terreno de las ideas como en la instrumentación práctica de las políticas que se llevan a cabo. Por ello, la evaluación de los programas, antes que los problemas de causalidad de las teorías, se deberá centrar sobre los procesos de instrumentación de las estrategias de intervención ya que son las contingencias que se establecen en cada caso las que —en última instancia— determinan las diferencias.

Este hecho ha supuesto que los enfoques relativos a la evaluación de programas desplazaran el foco de la reflexión hacia los problemas prácticos que conlleva toda estrategia de intervención ya que son los procesos los que, esencialmente, determinan la mayor o menor eficacia de los programas (Wholey, Hatry y Newcomer, 1994). De ahí que a la hora de evaluar sea muy importante distinguir entre los efectos que se deben a al desarrollo de los programas (*action hipotesis*) y los que están determinados por las teorías (*conceptual hipotesis*). Asumir esta distinción implica un cambio de enfoque sobre el concepto de programa ya que puede ser tipificado como una tecnología instrumental que debe ser diseñada y modificada de forma constante en función de su utilidad en relación con los objetivos establecidos.

Desde esta perspectiva los objetivos de la evaluación no se focalizan sobre la verificación de la inferencia causal de las teorías sino sobre las distintas hipótesis prácticas que se pueden formular en cada una de las fases que implica la conceptualización, el

291

diseño, la aplicación y la utilidad de una estrategia de intervención social. Por ello, un enfoque evaluativo global deberá abarcar objetivos relativos a cada una de estas fases, a saber: a) la estimación del mérito de las metas que se propone cada estrategia, b) las calidad del diseño y planificación de las actuaciones a realizar, c) el grado de cumplimiento y adecuación con que dicha planificación ha sido llevada a cabo (implementación) y d) la calidad y la utilidad de los resultados e impactos generados. La mayoría de los enfoques prácticos recientes coinciden en señalar estos cuatro objetivos como las metas a tener en cuenta en toda evaluación de un programa ya que nos permiten distinguir entre los fallos de la teoría y los fallos que se producen en la cadena de secuencias que implica el diseño y la aplicación de una estrategia de intervención social (Brinkerhoff, Brethower, Hluchyj y Nowakowski, 1985).

3.3. El programa como compromiso con el cambio social

En un tercer enfoque situaríamos aquellos que consideran que la justificación teórica de todo programa es promover un cambio social por lo que éste debe ser, igualmente, el objetivo de toda evaluación. Más allá de los problemas prácticos que conlleva toda estrategia de intervención, debemos asumir que su finalidad primordial es promover la transformación social. Por ello, en la medida que nuestra búsqueda persigue identificar las políticas, programas y acciones más adecuadas a este fin, el evaluador no puede realizar su trabajo sin cuestionar los valores y las consecuencias que se ponen en juego en cada estrategia de intervención. Lo cual significa que la calidad de su trabajo estará mediatizada no sólo por su capacidad técnica sino también por su compromiso en la resolución de los problemas sociales (pobreza, marginación, desigualdad, etc..) y en la lucha a favor de una sociedad más justa e igualitaria (Patton 1994, Fetterman 1994, De Miguel 1999).

La característica fundamental de este enfoque es que el evaluador no se sitúa al margen del programa sino que constituye uno de los elementos del mismo y, en esa medida, asume un papel crítico respecto al cambio social. Frente a los enfoques anteriores en los que desempeñaba el rol de experimentador y programador social, en este caso su trabajo se identifica con la misma finalidad que promueve el programa; es decir, necesita concebir su función dentro del conjunto de acciones que desarrolla el programa para promover el cambio y la mejora social (Nevo 1990, Sonnichsen 1994). Ahora bien, para asumir este papel es muy importante que el evaluador esté comprometido teórica y vitalmente en la resolución de los problemas sociales, situación que no siempre es un hecho.

Independientemente del nivel de aceptación de esta alternativa, lo que sí debemos reconocer es que ha contribuido decisivamente a resaltar dos aspectos realmente importantes en todo proceso evaluativo. De una parte, la utilidad de las evaluaciones debe estimarse en función de las mejoras sociales y la mayoría de las veces éstas no se producen porque no se toman las decisiones pertinentes que se derivan de los informes evaluativos. De otra, para que las evaluaciones tengan consecuencias, es preciso que los evaluadores se preocupen por maximizar el uso de los resultados de sus trabajos analizando y neutralizando, en la medida de lo posible, los factores políticos y con-

292

textuales que impiden que los sectores implicados tomen las decisiones oportunas. Cuando se asume el rol de agente de cambio, el evaluador no puede eludir ambos problemas, lo cual significa que no puede dar por concluido su trabajo mientras no sean tangibles las consecuencias; es decir, la mejoras generadas a partir de la evaluación.

Las diferencias entre las bases teóricas implícitas en estas tres formas de conceptualizar un programa social son evidentes. De ahí que existan muchas taxonomías que utilizan este criterio para clasificar los distintos enfoques teóricos y metodológicos que se aplican en los procesos evaluativos. Entre estas clasificaciones queremos resaltar dos que nos parecen muy pertinentes. Por un lado, la propuesta de Stufflebeam y Shinkfield (1987) *que agrupa los modelos en seudoevaluaciones, cuasievaluaciones y evaluaciones verdaderas* en función de la concepción y propósitos hacia los que se dirige el proceso evaluador. De otro, la que recientemente propone Chelimsky (1997) que clasifica los distintos enfoques teóricos en tres grandes perspectivas de acuerdo con la finalidad principal que orienta la evaluación: a) estimar los resultados y la eficiencia de los programas (*accountability perspective*), b) obtener conocimientos específicos sobre el programa como estrategia de intervención (*knowledge perspective*) y c) contribuir a mejorar el programa para incrementar la mejora y el desarrollo social (*development perspective*). Utilizando este criterio de clasificación, este último autor establece un contraste entre los posicionamientos de estas tres perspectivas respecto a nueve categorías o dimensiones que habitualmente utilizan para delimitar los supuestos de un enfoque teórico.

4. LA CONSTRUCCIÓN DEL CONOCIMIENTO

Independientemente del concepto que el evaluador tenga sobre el programa, la primera fase del proceso en toda evaluación es la *«búsqueda de información técnicamente apropiada para determinar el mérito o valor del objeto que vamos a evaluar»; es decir, la búsqueda de «evidencias» a partir de las cuales podamos justificar nuestras valoraciones.* Así pues, cualquiera que sea la finalidad que oriente el proceso, el primer objetivo de una evaluación es obtener informaciones pertinentes, válidas y fiables que actúen como soporte para nuestros juicios. Ello implica que, en esta búsqueda, debamos conjugar dos tipos de criterios —utilidad y rigor— que no siempre se tienen en cuenta. Para ello, el evaluador deberá reflexionar sobre el tipo de conocimientos que necesita sobre el programa y cómo puede proceder para su construcción; es decir, concretar a qué denomina «información técnicamente apropiada» y, en consecuencia, cómo va a proceder para su obtención e interpretación. Ambas cuestiones están estrechamente relacionadas por lo que las ideas previas que tiene el evaluador sobre las cuestiones epistemológicas serán determinantes a la hora de tomar estas decisiones.

4.1. Bases teóricas del conocimiento

Dado que existen diversos tipos de conocimientos (causación, generalización, predicción, descripción, comprensión, etc.), el evaluador se verá obligado a justificar cuá-

les van a constituir el objeto de su búsqueda ya que, en cada trabajo, unos pueden ser más adecuados que otros. La predilección por un tipo específico de metodología sólo puede ser justificada a partir de los supuestos que fundamentan toda teoría del conocimiento; es decir, las bases ontológicas, epistemológicas y metodológicas que sustentan la evaluación de programas como una disciplina inserta en el marco de la investigación social empírica. En la práctica es habitual especificar los planteamientos metodológicos con que se aborda la evaluación de un programa pero rara vez se hace mención alguna a las concepciones sobre la naturaleza de la realidad que es objeto de análisis (bases ontológicas) o sobre los fundamentos para la construcción del conocimiento científico (bases epistemológicas). Los evaluadores normalmente consideran que abordar estas cuestiones es algo tangencial a su trabajo por lo que suelen evitarlas.

Esta postura, aunque es comprensible, es realmente engañosa ya que —como señalan Shadish, Cook y Leviton (1991)— *la opción metodológica elegida por un evaluador implica implicitamente una toma de postura ontológica y epistemológica.* Por ello, a efectos de que no se produzcan graves disfunciones entre lo que se piensa sobre una realidad (programa) y lo que se hace en la práctica para conocerla, es necesario que los evaluadores se habitúen a explicitar las bases teóricas que justifican sus planteamientos metodológicos; es decir, a conectar los métodos que utilizan con los parámetros teóricos que consideran pertinentes para la construcción del tipo de conocimiento —explícito o tácito— que esperan obtener a través del proceso evaluativo.

En la práctica esta situación se resuelve acudiendo a los paradigmas clásicos que se emplean en la investigación social, por lo que habitualmente se utilizan los distintos supuestos o principios que definen cada uno de ellos para justificar los distintos enfoques u modelos evaluativos. Ello determina que con frecuencia se hable de los paradigmas positivista, naturalista y crítico —u otras de las denominaciones al uso— aplicados al campo de la evaluación de programas sin previamente reconceptualizar la idoneidad de los mismos para el tipo de conocimiento que se persigue en cada caso y, en consecuencia, sobre los métodos a utilizar. Entendemos que esta extrapolación constituye un apriorismo injustificado, ya que, de una parte, no tiene en cuenta que en este campo no existe ningún paradigma que resulte intachable —ya que los problemas y dificultades son cuestiones inherentes a todos ellos— y, de otra, supone abordar el tema de la diversidad paradigmática desde las tesis de la incompatibilidad o complementariedad, al margen de las tendencias actuales sobre la unidad (Walker y Evers 1988, Howe 1992).

Precisamente, debido a que ninguno de los paradigmas clásicos ofrece unas bases epistemológicas plenamente satisfactorias al tipo de conocimiento que requiere esta realidad, diversos autores tratan de promover un marco teórico adecuado que justifique esta disciplina que cada día adquiere más relieve. Entre los esfuerzos realizados en esta línea, consideramos muy interesante la propuesta reciente de un «nuevo marco paradigmático» orientado hacia la explicación social realista (Pawson y Tilley 1997) cuyos supuestos teóricos se enmarcan en la tradición que sitúa la explicación científica entre el positivismo y el relativismo (Hesse 1974, Blaskar 1975, Sayer 1984). Avanzar en esta línea nos puede aportar un marco epistemológico más adecuado que acudir a los paradigmas clásicos.

4.2. Naturaleza de las evidencias: datos y opiniones

Ubicados en el plano del realismo, hemos apuntado que el primer objetivo de todo proceso evaluador es la obtención de «evidencias» sobre las que podamos apoyar nuestros juicios. En sentido etimológico se dice que se tiene evidencia de algo cuando se tiene certeza; es decir, *«cuando la información que disponemos nos ofrece certidumbre manifiesta y perceptible»*. Ahora bien, en los procesos evaluativos recogemos informaciones de distinta naturaleza y no todas nos ofrecen el mismo grado de certidumbre. En unos casos son datos —cuantitativos y cualitativos— obtenidos a través de procedimientos sistemáticos —más o menos rigurosos— que nos aportan informaciones precisas sobre diversos aspectos del programa. En otros recabamos las opiniones (percepciones, expectativas, valoraciones, etc...) que nos manifiestan el concepto o juicio que formulan sobre diversos aspectos del programa las audiencias implicadas en el mismo.

El problema radica en a qué tipo de información le atribuimos la categoría de conocimiento cierto, válido y fiable. Algunos evaluadores, inmersos en la lógica positivista, sólo aceptan como certezas aquellos datos «objetivos» que han sido obtenidos mediante la utilización de una metodología científica rigurosa. Este apriorismo les lleva lógicamente a centrar toda búsqueda de información sobre procedimientos metodológicos «duros» en los que se garantiza la validez interna y externa de los datos obtenidos. Digamos que prevalecen los criterios formales a los prácticos, independientemente de su adecuación para la evaluación del diseño, desarrollo y mejora de un programa y de su utilidad para las audiencias implicadas. Con frecuencia un tratamiento y análisis sofisticado de los datos impide que los resultados y los informes de una evaluación sean comprensibles para aquellos que deben tomar decisiones (Cook y Reichardt 1986, Fetterman y Pitman 1986).

Frente a ellos situaríamos a quienes presentan una actitud más pragmática al considerar que el rango de un conocimiento viene determinado por su utilidad, lo cual significa ante todo preguntarse por el tipo de información que se precisa para evaluar cada fase de un programa y, especialmente, la que necesitan los clientes y las audiencias implicadas para tomar decisiones. Cuando se adopta este punto de vista resulta imprescindible recabar «otro tipo de conocimientos» construidos fundamentalmente a partir de datos «subjetivos» y hacerlo, igualmente, con la adecuada metodología sistemática para que puedan ser considerados como evidencias contrastadas. *De hecho muchos de los trabajos evaluativos a realizar tienen esta finalidad: estimar el parecer o juicio sobre el programa que comparten la generalidad de personas implicadas en el mismo (opinión pública).*

En definitiva, el evaluador deberá precisar a qué tipo de conocimientos les atribuye la categoría de «evidencias». Todos sabemos que existen muchas clases diferentes de conocimientos que pueden ser objeto de nuestra búsqueda sólo que, en cada estudio y fase del proceso evaluativo, se pone el énfasis sobre un determinado tipo o categoría en función de su pertinencia y oportunidad. Ello no significa que unos sean mejores que otros, ya que todos tienen sus ventajas e inconvenientes. Lo aconsejable es abordar la evaluación de un programa integrando distintas informaciones a partir de estudios

y métodos diversos que nos permitan obtener conocimientos más completos, fiables y válidos, porque «*la calidad de un conocimiento se incrementa en la medida que existe un mayor escrutinio público sobre el mismo*» (Shadish, Cook y Leviton, 1991).

4.3. La controversia entre deducción e inducción

Partiendo de las consideraciones anteriores fácilmente se desprende que tan oportunos y adecuados son los procedimientos metodológicos deductivos como los inductivos para la obtención de «evidencias». En los casos en los que buscamos comprobar relaciones entre inputs y outputs u obtener conocimientos generalizables es lógico que procedamos con una metodología deductiva rigurosa que nos facilite una información conclusiva al respecto. Por el contrario, si el propósito no es generalizar una estrategia sino analizar y comprender lo que sucede en cada intervención para contribuir a mejorar el programa, entonces lógicamente resultará más adecuado llegar al conocimiento por vía inductiva a partir de las distintas informaciones e interpretaciones que nos aporten los implicados. No cabe, por tanto, contraponer dos tipos de procedimientos —verificación frente a descubrimiento— porque buscan cosas diferentes.

De ahí que los problemas teóricos que conlleva la construcción del conocimiento en un campo tan complejo como es la evaluación no se resuelvan señalando que se opta por tal o cual metodología cuantitativa o cualitativa. La metodología a utilizar es la resultante de una serie de decisiones teóricas que previamente hay que definir y precisar de forma explícita si queremos justificar nuestros procedimientos. Por ello, una vez que hemos determinado qué clase de conocimientos se necesitan y por qué, será el momento de proceder a concretar las bases teóricas que justifican nuestros métodos. Ello nos permitirá discernir que no todos los métodos son aconsejables para un mismo objetivo y que es muy importante considerar las ventajas e inconvenientes de cada uno de ellos en función de los diferentes propósitos que pretendemos alcanzar con una evaluación. Lo cual significa que, desde el punto de vista epistemológico, no existe ninguna primacía de unos métodos sobre otros. De hecho, *una de las principales características de la evaluación de programas es que sus métodos cubren toda la gama de paradigmas que dominan en la investigación social* (Rossi, Freeman y Lipsey, 1999).

Si se asume esta postura, la consecuencia lógica es que no cabe enfrentar o contraponer las metodologías cuantitativas y cualitativas como más o menos idóneas para la evaluación de programas ya que todas son necesarias. Las características específicas de los programas y de los procesos que conlleva su conceptualización, diseño, implementación y evaluación nos obliga a considerar fenónemos de naturaleza distinta y en contextos diversos para cuyo conocimiento se requieren métodos diferentes en cada caso, por lo que la polémica respecto a los paradigmas metodológicos queda en un segundo plano (Cook y Reichardt 1986, Fetterman, Kaftarian y Wandersman, 1996). De ahí que, *la evaluación de programas constituya actualmente el ámbito de la investigación social aplicada en el que se hace más patente la superación del debate entre metodología cuantitativa y cualitativa, lo cual significa que clasificar los modelos evaluativos según este criterio carece de sentido* (Reichart y Rallis 1994, Cook 1997, De Miguel 1999).

5. EL PROBLEMA DE LOS CRITERIOS DE VALOR

La evaluación de un programa no se limita a una mera recolección, análisis e interpretación de los datos para obtener «evidencias». Una vez que disponemos de la información necesaria comienza la siguiente fase del proceso: formular los juicios de valor. A la vista de los datos obtenidos deberemos emitir nuestra valoración de forma clara, precisando el rango y matices de los juicios (ausencia o presencia de calidad; puntos fuertes y débiles, recomendaciones y propuestas, etc..). Ahora bien, *para proceder a su formulación el evaluador se encuentra con otro problema: concretar los criterios o marcos de referencia que va a utilizar para emitir sus juicios.* Los datos obtenidos sobre un programa son simplemente datos —más o menos probados— pero su valoración depende del criterio elegido. Precisamente, debido a que en la práctica no existen evaluaciones libres de valores, el evaluador deberá especificar en cada caso los criterios que utiliza como punto de referencia para formular sus juicios respecto a los distintos aspectos del programa.

Habitualmente en la literatura se hace la distinción entre los conceptos de mérito y valor como dos marcos de referencia distintos —intrínseco y extrínseco— a la hora de formular los juicios (Guba Lincoln, 1981, 1989, Scriven 1991), términos que a su vez nos remiten a otra clasificación de los criterios de evaluación en dos grandes bloques: criterios descriptivos y criterios prescriptivos (Alkin y Ellet, 1990). Realizamos una valoración descriptiva cuando se elige o asume «valores de otros» como criterio de referencia para emitir los juicios sobre el programa. Por el contrario, denominamos una valoración como prescriptiva cuando es el propio evaluador el que define los criterios que considera adecuados y correctos para emitir los juicios, al margen de que éstos sean o no coincidentes con los que defienden otras audiencias.

Las diferencias entre ambos enfoques se perciben fácilmente. Cuando realiza una valoración decriptiva nuestro papel se limita a preguntarles a los demás los criterios a utilizar y aplicarlos con rigor y corrección. En cambio, optar por una valoración prescriptiva exige decantarse por unos criterios frente a otros lo que supone previamente adoptar una postura personal ante todas las cuestiones relacionadas con los valores que se ponen en juego cada programa. En el primer caso, al aceptar los objetivos de los demás como criterios de valoración, el evaluador puede conservar su independencia y mantenerse al margen de intereses sectoriales. En el segundo, dado que no existe una teoría general que pueda considerarse correcta sobre la valoración prescriptiva, es muy problemático que un evaluador —especialmente cuando actúa como externo— utilice sus propios criterios como marcos de referencia para evaluar un programa y que estos sean aceptados por todas las partes (Vedung, 1997).

Por esta razón, en la práctica de las evaluaciones relativas a programas sociales es más común utilizar criterios descriptivos que prescriptivos. Aunque estos últimos, en tanto que defienden la primacía de valores absolutos (igualdad, justicia, calidad intrínseca, etc...) cumplen una función muy importante a lo largo de todo el proceso evaluativo, son realmente las teorías descriptivas las que nos proporcionan criterios más legítimos y útiles tanto desde el punto de vista social como técnico (Shadish, Cook y Leviton, 1991). De ahí que la mayoría de los evaluadores defiendan la utilización de este tipo de criterios para emitir los juicios de valor sobre los programas.

No obstante, el problema no concluye al señalar que se asume como referencia una valoración descriptiva ya que los «otros» no son un conjunto homogéneo sino que tienen, o pueden tener, intereses diversos y, en ocasiones, contrapuestos. Necesitamos, por tanto, señalar cuál es la perspectiva valorativa externa que utilizamos para emitir los juicios; es decir, especificar el criterio que elegimos para juzgar que un programa es «bueno». En este sentido cabe considerar que existen tres grandes opciones para el evaluador según utilice preferentemente criterios políticos, democráticos o técnicos.

5.1. Criterios políticos

Todos conocemos que la mayoría de los planes y programas de intervención social están patrocinados por organismos y servicios públicos. Ello significa que tanto su concepción como el diseño y la aplicación normalmente responden a determinados supuestos ideológicos e intereses de la política que los sustenta. Los programas constituyen formalmente la concreción de unas decisiones políticas respecto a unas prioridades en materia de actuación social, lo que indudablemente supone apostar por un tipo de valores. Esta apuesta no sólo se traduce en promover un determinado tipo de programas sino también en todas las decisiones que implica la selección y priorización de estas estrategias, especialmente las relativas a la redistribución y asignación de los recursos.

Cabe, por tanto, hablar de un tipo de evaluación que asume como criterios de valoración los objetivos —políticos, económicos, sociales, etc..— que se proponen alcanzar los organismos y agencias que patrocinan los diversos programas. El evaluador, en estos casos, acepta los valores del patrocinador (gestores, administradores, empresarios, etc...) como punto de referencia para emitir sus juicios y así poder facilitarle la información que necesita para que pueda tomar las decisiones pertinentes. Este hecho no implica en modo alguno que el evaluador, tanto si centra su tarea sobre los objetivos cubiertos como sobre los recursos empleados, pierda su objetividad convirtiendo su trabajo en una tarea burocrática al servicio de un estamento político. Al contrario, sus análisis y recomendaciones pueden cuestionar tanto los criterios que inspiran las políticas públicas como las acciones desarrolladas y, en consecuencia, promover cambios internos importantes.

5.2. Criterios democráticos

Frente a la utilización exclusiva de los valores de los «estamentos políticos», cabe pensar que quienes mejor conocen la oportunidad y efectividad de toda estrategia de intervención social son la personas directamente implicadas en la misma por lo que sus criterios deben constituir un marco de referencia preferente para formular los juicios de valor. Al margen de los planteamientos de los gestores, debemos considerar que todas las personas que constituyen las audiencias de un programa generan —en menor o mayor grado— una serie de expectativas, percepciones y valoraciones intrínsecas que es necesario tener en cuenta a la hora de emitir las valoraciones sobre su oportunidad y eficacia. Más aún, en tanto que es algo que les afecta directamente, no

sólo tienen derecho a ello sino que no podemos prescindir de sus criterios para valorar el programa (De Miguel, 1995). Cuando el evaluador asume este marco de referencia decimos que realiza una valoración descriptiva pluralista basada en criterios democráticos.

Ahora bien, para que el evaluador asuma este tipo de perspectiva valorativa es necesario que previamente haya enfocado todo el proceso evaluativo de forma participativa; es decir, teniendo en cuenta que quienes están directamente implicados en un programa son los que deben ser los actores principales de su evaluación. Los argumentos para ello están claros. De una parte, son los que conocen más de cerca la realidad del mismo y, por tanto, pueden facilitarnos informaciones más pertinentes e interesantes. De otra, son los más implicados en las consecuencias y difícilmente se podrán poner en marcha mejoras si ellos no participan activamente en su valoración y elaboración. Todas las propuestas y postulados actuales sobre evaluación pluralista insisten en esta línea: *la oportunidad y conveniencia de utilizar criterios democráticos a la hora de diseñar, aplicar y evaluar cualquier estrategia de intervención social* (Fetterman, Kaftarian y Wandersman 1996, Floc'Hlay y Plottu 1998, Apodaca 1999).

5.3. Criterios técnicos

En una posición intermedia entre la política y la ciudadanía situaríamos la valoración técnica; es decir, aquella que lleva a cabo el evaluador mediante la aplicación de procedimientos e instrumentos científicos apropiados. Los procesos de evaluación conllevan la aplicación de una serie de reglas y principios sobre estrategias de intervención social cuya correcta utilización contribuye a clarificar tanto la adecuación de los programas como la calidad de sus resultados (Martin 1993, Martin y Kettner, 1996). Cuando el evaluador asume este enfoque utiliza la propia teoría normativa sobre evaluación de programas como criterio o marco de referencia a la hora de formular sus juicios de valor; es decir, la calidad de las metas propuestas, la coherencia del diseño efectuado, la adecuación de los procesos de implementación, la efectividad de los resultados e impactos generados y la oportunidad y viabilidad de las decisiones tomadas en relación con los procesos de mejora.

No obstante, el problema no se resuelve señalando que se van a utilizar criterios técnicos ya que supone dejar los marcos de referencia demasiado abiertos. De ahí que el evaluador deba precisar también las normas, estándares e indicadores que utilizará en cada trabajo para valorar los diferentes aspectos del programa. En unos casos para establecer estas normas podrá acudir a referencias nacionales e internacionales al uso (estándares de calidad, normas internacionales, protocolos sobre buenas prácticas, etc..). En otros se verá obligado a establecer criterios y elaborar indicadores que le permitan objetivar el proceso de valoración y justificar ante las audiencias los juicios emitidos. La dificultad que entraña esta tarea determina que sea una de las cuestiones técnicas más problemáticas y controvertidas en todos los procesos evaluativos (De Miguel, 1991).

Dado que cabe utilizar diferentes perspectivas valorativas como marco de referencia para formular los juicios, se pueden igualmente establecer taxonomías evaluativas

utilizando este criterio de clasificación. Así, algunos autores hablan de *modelos descriptivos y modelos prescriptivos* (Alkin y Ellet, 1990), y otros de *evaluación burocrática, autocrática y democrática* (Mac Donald, 1977). También se distingue entre *modelos preordenados o convencionales y modelos pluralistas o participativos* según el grado de implicación de las audiencias en los procesos de valoración (Lincoln y Guba 1985, Guba y Lincoln 1989, Rebien 1997). Ahora bien, independientemente de la oportunidad y adecuación de estas clasificaciones, lo que sí debemos constatar es que el cuestionamiento actual de criterios políticos y técnicos ha generado una fuerte tendencia a favor de los criterios democráticos que se manifiesta en todos los enfoques evaluativos actuales —*evaluación pluralista, evaluación participativa, evaluación respondente, evaluación democrática, cluster evaluation, empowerment evaluation*, etc.— cuyo denominador común se caracteriza por concebir el trabajo del evaluador como un compromiso con los derechos de todas las partes con el fin de contribuir a la educación cívica, el desarrollo humano y la emancipación social (Fetterman 1994, Smith 1998).

6. UTILIDAD Y CREDIBILIDAD DE LOS INFORMES

Toda evaluación debe concluir con la toma de decisiones que se deriven de las valoraciones formuladas. Aunque el objetivo inmediato de todo proceso es ofrecer juicios razonados para la toma de decisiones, la finalidad última de toda evaluación es lograr que tenga consecuencias. Precisamente, porque todos esperamos que nuestro trabajo sea útil para resolver problemas sociales, el evaluador no pueda prescindir del análisis y explicación de las relaciones que existen entre los resultados y recomendaciones que emite en sus informes, las decisiones que se toman al respecto y las consecuencias que éstas tienen socialmente. No preocuparse por estas cuestiones pone en tela de juicio la credibilidad de nuestro trabajo. Por ello, la reflexión teórica que conlleva el diseño y aplicación de una estrategia evaluativa debe abarcar también el análisis de los factores —cómo, cuándo, dónde, para qué—, que condicionan la utilización posterior de los resultados que se aportan en los informes (Rebien 1997, Chelimsky 1997).

Ahora bien, abordar el tema de cómo se puede maximizar el uso de los informes evaluativos (*knowledge use*) no una tarea fácil ya que son múltiples los problemas y factores implicados. No obstante, cabría agrupar esta problemática en torno a cuatro preguntas o interrogantes fundamentales: ¿Qué clases o tipos de usos pueden tener nuestros informes?, ¿Quiénes son los usuarios potenciales de los mismos?, ¿Qué papel puede jugar el informe para estos usuarios?, ¿Cuáles son los factores políticos y contextuales que pueden dificultar la utilización de los resultados? Efectuaremos algunos comentarios sobre estas cuatro cuestiones.

6.1. Posibles usos del Informe

Abordar el tema de la utilidad o usos del informe nos remite nuevamente a la finalidad que promueve y orienta todo el trabajo evaluativo, ya que entre ambas cuestiones existe una mutua implicación. Como ya hemos comentado, existen propuestas

diferentes en relación a los fines de una evaluación aunque, tomando como referencia los usos, se podrían agrupar las diversas alternativas en dos bloques: a) los que entienden que la finalidad primordial de una evaluación es la búsqueda de «observaciones y explicaciones generalizables» mediante la utilización de procedimientos científicos rigurosos y b) los que consideran que lo verdaderamente importante es facilitar a todos los implicados la información que necesitan para tomar decisiones (stakeholders services). En el primer caso se conciben el informe evaluativo de forma similar al de una investigación por lo que los problemas de difusión, diseminación y utilización de los conocimientos son muy afines a los que suceden en otras disciplinas sociales y que se conocen a través de la literatura especializada (Keeves, 1988). En el segundo, dado que cada evaluación de un programa es una tarea singular, el evaluador se ve obligado plantearse en cada proyecto cuál puede ser la utilización posterior o usos posibles que puede tener la información que se aporta en los informes. Al respecto, las demandas y usos más frecuentes que se hacen de los informes evaluativos son los siguientes: *para tomar decisiones, identificar problemas, inducir acciones y justificar decisiones* (Braskamp y Brown 1980, Ewell 1989).

6.2. Identificar los usuarios potenciales de la información

Precisamente porque entendemos que la finalidad prioritaria en toda evaluación de programas es facilitar a las audiencias la información que necesitan, consideramos que, inicialmente, el evaluador deberá identificar quiénes son los usuarios potenciales de su trabajo ya que de ello dependen los tipos de usos que pueden tener sus informes. Deberá tener en cuenta, además, que aunque los trabajos sean objeto de contrato no significa en modo alguno que los informes tengan utilidad solamente para los clientes. De ahí que una de las primeras cuestiones para enfatizar la utilización de los resultados sea la identificación de los usuarios potenciales de nuestro trabajo con el fin de ofrecerles la información que necesitan, lo que implica, previamente, analizar los problemas específicos que se dan en cada contexto y programa: audiencias implicadas, las posibilidades que tienen de acceso a la información, tipos de informes a elaborar, etc.. En toda evaluación existen muchos «implicados» que, en mayor o menor medida, pueden hacer uso de los informes siempre que estos les ofrezcan informaciones oportunas y asequibles Aunque algunos estén más comprometidos que otros en los procesos de toma de decisiones ello no significa que los trabajos no les sean útiles. En la medida que hacemos llegar informes adecuados a todas las partes implicadas en un programa —gestores, administradores, agencias, instituciones, científicos, técnicos, implicados, ciudadanos, etc.— incrementamos las posibilidades de que nuestro trabajo tenga mas consecuencias.

6.3. Clarificar el nivel de operatividad del informe

Además de identificar los usuarios potenciales de nuestro trabajo también tenemos que plantearnos el papel que va jugar el propio informe en los procesos de toma de decisiones. Aunque existen muchas propuestas al respecto, siguiendo a Shadish, Cook

y Leviton (1991), *cabe establecer tres tipos de planteamientos donde los informes se utilizan con criterios diferentes: instrumental, conceptual y persuasivo.* Se considera que un informe es instrumental cuando los resultados obtenidos se utilizan de forma directa para tomar decisiones sobre el programa. Se considera como informe conceptual aquel cuyas valoraciones no suelen ser utilizadas como soporte para decidir cambios pero que, indirectamente, influyen sobre cómo perciben y valoran las personas las intervenciones sociales. Finalmente se entiende que el informe cumple una función persuasiva cuando las evaluaciones son utilizadas para convencer a las audiencias que tienen estimaciones apriorísticas sobre los programas de que la evaluación es el proceso correcto para valorar cualquier estrategia de intervención. Precisar estas cuestiones previamente no sólo nos ayudará a elaborar informes más operativos y útiles desde el punto de vista práctico sino también a conceptualizar la función o papel que le otorgan a nuestro trabajo las partes implicadas en una evaluación. Con el fin de evitar confusiones, algunas de estas cuestiones deben quedar explícitas en el protocolo que se elabore en la fase de negociación de proyecto.

6.4. Elaboración y difusión adecuada de los informes

Finalmente, para maximizar el uso de los informes también debemos considerar otros problemas que tienen que ver con el proceso de elaboración y difusión de los resultados. Respecto a la elaboración, el evaluador deberá preocuparse de facilitar a cada audiencia la información precisa, con datos que pueda fácilmente interpretar y con un lenguaje adecuado. De ahí que en este campo no sea oportuno utilizar técnicas de tratamiento y análisis de datos complejas cuyos resultados son difíciles de valorar salvo para especialistas. Los problemas de la difusión generalmente depende de factores de orden político que se generan en cada contexto de forma especial. En ocasiones, quienes encargan la evaluación se adueñan del informe y sólo lo difunden en la medida que beneficia a sus intereses. En el sentido opuesto situaríamos los planteamientos que consideran que todos los informes deben ser difundidos en términos inteligibles para todas la audiencias con el fin de que todos puedan participar y valorar tanto la utilización que se hace de los resultados como de las decisiones que se toman al respecto. Aunque las posibilidades del evaluador al respecto son bastante limitadas, ello no significa que no deba tener en cuenta estos problemas (McLaughlin y McLaughlin 1989, Wholey, Hatry y Newcomer 1994).

En resumen, aunque inicialmente la utilidad de una evaluación depende de los propósitos, audiencias e intereses en función de los cuales se establece, en la práctica son los procesos de toma decisiones los que determinan finalmente las consecuencias que tiene cada proceso. Por ello, la tarea del evaluador no concluye con la elaboración del informe sino que debe reflexionar sobre los problemas que genera su utilización para que tenga consecuencias. Asumir esta perspectiva implica cuestionar no sólo qué información necesitan quienes deben tomar decisiones (evaluación orientada a la decisión) sino también la que deben conocer todos los implicados (evaluación orientada al usuario). No es lo mismo pensar que las decisiones dependen de los gestores y administra-

dores de los programas que involucrar también a las audiencias, clientes e implicados. Tampoco es igual que el informe aporte datos concluyentes en términos técnicos o se faciliten informaciones y recomendaciones adecuadas a las diversas audiencias. Fácilmente se advierte que la utilidad de una evaluación es mayor en la medida que tienen posibilidad de acceso a la información todos los sectores implicados.

De lo cual se deduce que el problema de la utilidad y usos de los informes es una cuestión compleja que depende de factores diversos (identificación de usuarios potenciales de los informes, claridad de los resultados, análisis de las alternativas, procesos de toma de decisiones, etc..), que necesitan ser analizados con cierto detalle para evitar obstáculos que impiden su utilización. De ahí que, durante los últimos años, hayan aparecido diversas corrientes y enfoques metodológicos que centran su preocupación sobre las consecuencias de los informes: *decision-oriented evaluation* (Stufflebeam y al., 1971), *utilization-focused evaluation* (Patton, 1978) *stakeholder-based evaluation* (Bryk, 1983), *beyond decision-oriented evaluation* (Sheerens, 1987), *outcomes-based evaluation* (Schalock, 1995), etc.. Aunque estos enfoques presentan entre ellos matices distintos, el denominador común de todos ellos es la necesidad de abordar los procesos evaluativos teniendo en cuenta que la credibilidad del evaluador se halla comprometida por la utilidad y usos que se derivan de sus informes.

7. DELIMITACIÓN DEL ENFOQUE METODOLÓGICO

Al iniciar el presente trabajo hemos distinguido entre fundamentos y enfoques evaluativos como dos términos relativos a la evaluación del programa que tienen un significado diferente. También hemos dicho que los evaluadores suelen ocuparse más de precisar los segundos que los primeros por lo que hemos considerado oportuno abordar algunos conceptos relacionados con los fundamentos porque, desde nuestro punto de vista, éstos deben ser objeto de reflexión antes de optar por un determinado enfoque metodológico. La elección de una aproximación o enfoque específico no se limita a una simple distinción entre metodología cuantitativa y cualitativa ya que supondría una simplificación absurda de una realidad compleja que tiene muchos matices y requiere ser abordada de forma comprensiva.

En la práctica, la determinación de un enfoque metodológico (*approach*) se conoce como la elección del modelo y el diseño de una evaluación. Cuando hablamos de «modelo» hacemos referencia a la forma o modo de abordar y representar la conceptualización teórica del programa que consideramos oportuna para llevar a cabo su evaluación una vez que hemos analizado los supuestos o fundamentos teóricos que justifican esta actividad: naturaleza del programa, la finalidad que nos proponemos, audiencias implicadas, información necesaria, metodología a utilizar, criterios de evaluación, utilidad o usos de los resultados, etc. Por el contrario, el término «diseño» alude al proceso que deberemos seguir para obtener las informaciones que reúnan la categoría de «certezas». El objetivo de los modelos es lograr una representación teórica sobre los fenómenos a investigar y los procesos a realizar que sea coherente y adecuada a la naturaleza del programa y los propósitos de nuestro trabajo. La finalidad de los diseños es suministrar a los investigadores una lógica científica que aporte validez tanto a los

procesos de evaluación como a los resultados. Así pues, mientras que el modelo alude al marco teórico que justifica el enfoque que hemos elegido para evaluar un programa, el diseño hace referencia a la estrategia metodológica que vamos a utilizar para que los datos obtenidos puedan ser considerados como evidencias (De Miguel, 2000).

7.1. La elección del modelo evaluativo

Al igual que sucede en toda investigación social, la elección de un modelo para evaluar un programa implica tomar decisiones sobre una serie de cuestiones teóricas relativas al tipo de estrategia —programa— que se pretende evaluar, las bases epistemológicas que justifican la construcción de los conocimientos, los criterios de valoración y la finalidad o usos que justifican la evaluación. Las preguntas claves que se debe formular un evaluador antes de decidir el marco teórico en el que va a situar su trabajo son las siguientes: ¿Qué quiero saber o qué quiere saber el cliente que me encarga la evaluación en relación con el programa? ¿De qué tipo de programa estamos hablando?, ¿Qué pretende hacer y/o decidir el cliente respecto al programa objeto de estudio? ¿Qué tipo de información necesito? ¿Cuáles son los procedimientos idóneos para su obtención? ¿Cuál es el papel que voy a desempeñar?, etc... Las respuestas a estas cuestiones son las que realmente determinan el enfoque o aproximación metodológica y, en consecuencia, las tácticas y estrategias que justifican un proceso evaluativo que debemos adoptar.

7.1.1. Finalidad del estudio evaluativo

El primer paso para elegir un modelo evaluativo es identificar los fines que se pretenden alcanzar. La identificación de los fines se puede realizar mediante dos procedimientos: según el tipo de función que persiguen los estudios y de acuerdo con los objetivos concretos que se plantean. Respecto a lo primero, actualmente se considera que los procesos evaluativos se orientan a una de estas tres funciones básicas: la rendición de cuentas (*accountability*), la mejora de los programas (*development*) y el incremento de conocimientos respecto al mismo (*knowledge*) (Chelimsky, 1997). Desde el punto de vista de los objetivos los principales propósitos que determinan la petición de procesos evaluativos se pueden agrupar en tres grandes bloques, a saber: a) Comprobar si los objetivos de los programas han sido alcanzados, b) Recabar opiniones fundamentadas sobre la valoración que hacen del programa las audiencias implicadas, y c) Obtener informaciones cualificadas a partir de las cuales se pueda tomar decisiones sobre el futuro de un programa específico (Vedung 1997). A estas tres alternativas se añade una cuarta cuando el objetivo es estimar la relación entre los recursos empleados y los resultados obtenidos (coste-efectividad, coste-beneficio) que da lugar a los modelos económicos (Lewin 1983, Drummond, Stoddart, y Torrance 1991, Nas 1996).

El evaluador deberá precisar cuál o cuáles son los propósitos concretos que realmente promueven cada trabajo con el fin de adoptar un marco teórico adecuado. Aunque ciertamente cada persona puede generar una comprensión «peculiar» de los intereses de los patrocinadores y las necesidades de los clientes y, en consecuencia,

establecer con libertad los parámetros teóricos que considera apropiados para enfocar el proceso evaluativo, procede señalar que «el modelo a utilizar depende del tipo de información que se solicita y que el evaluador debe ajustar su elección a la finalidad que justifica el encargo». En otros términos, los modelos deben ser elegidos en función de los propósitos del estudio.

7.1.2. Concreción de los parámetros del modelo

Una vez que hemos concretado los fines procede delimitar el resto de dimensiones, factores y variables que intervienen en la determinación y representación de un modelo. Como ya hemos comentado en otro momento (De Miguel, 1989), para que un modelo pueda servir de guía, además de estar fundamentado sobre una teoría, necesita concretar los fenómenos que intervienen y las relaciones que se establecen entre los mismos. Exige, además, especificar operacionalmente las variables con el fin de probar la validez empírica del modelo y la posibilidad de generalizar su aplicación. En el caso de la evaluación de un programa conlleva, además, una toma de postura sobre el papel de los valores y rol que el evaluador va a desempeñar ante todos los factores implicados en el proceso a los que hemos aludido anteriormente.

Así pues, una vez clarificada la finalidad y antes de tomar una opción metodológica, el evaluador deberá reflexionar y explicitar los distintos supuestos o fundamentos que adopta en cada caso concreto y que constituyen los determinantes inmediatos en la elección de modelo. A la vista de la literatura (Shadish, Cook y Leviton 1991, Rebien 1997), no existe consenso entre los autores sobre cuáles deben ser los factores o dimensiones que se consideran básicas para la construcción de un modelo evaluativo, como paso previo para su especificación. Nuestra propuesta al respecto es que, al menos, se deben considerar los parámetros que hemos abordado anteriormente, a saber: *la conceptualización del programa como estrategia de intervención social, las bases epistemológicas que determinan la construcción de los conocimientos, el tipo de criterios a utilizar para la formular los juicios de valor, la utilidad o usos de que pueden tener los informes y el papel a desempeñar por el evaluador a lo largo del proceso.* Consideramos que estas dimensiones aglutinan las variables o características más importantes a partir de las cuales se puede definir un modelo evaluativo

7.1.3. Utilización de las taxonomías

Generalmente los evaluadores no construyen un modelo específico para cada evaluación sino que utilizan como marco de referencia alguno de los que ya están formalizados en algunas de las taxonomías al uso (MacDonald 1977, House 1980, Guba y Lincoln 1981, Madaus, Scriven y Stufflebeam 1983, Worthen y Sanders 1987, Stufflebeam y Shinkfield 1987, Stake y Mabry 1997, etc...). El problema que presentan estas taxonomías es que están elaboradas con distintos criterios de clasificación que impiden cotejar y contrastar las distintas categorías que establecen. La elección de una u otra como marco de referencia ya implica una toma de postura del evaluador que si no se hace de forma razonada puede, en principio, generar incoherencias al no existir

305

correspondencia entre los supuestos teóricos del modelo elegido y el tipo de evaluación que procede realizar en función de la finalidad.

Entre estas taxonomías nos interesa destacar de forma especial la propuesta por Stecher y Davis (1991) ya que —desde nuestro punto de vista— nos parece realmente útil pues no sólo tiene en cuenta como criterio de clasificación la finalidad de la evaluación sino que también utiliza para su elaboración dimensiones muy semejantes a las que nosotros proponemos: la naturaleza del programa, las preguntas o cuestiones básicas a resolver, la información que necesitamos, el eje o núcleo sobre el que se centra el proceso y el rol que el evaluador desempeña en el proceso. A partir de estos criterios, *estos autores agrupan los modelos evaluativos en cinco enfoques —experimental, dirigido a metas, orientado a las decisiones, centrados sobre los clientes y respondente— que pueden ser considerados como arquetipos representativos de las diversas opciones o aproximaciones metodológicas que se pueden plantear a la hora de abordar la evaluación de un programa.* Desde nuestro punto de vista, esta clasificación de los modelos nos parece realmente interesante y útil como ya hemos comentado en otro momento (De Miguel, 2000).

7.2. La elección del diseño y los métodos

Una vez que hemos decidido el enfoque o modelo que vamos a asumir para llevar a cabo la evaluación procede establecer el diseño; es decir, «la estrategia o plan que vamos a seguir donde se especifica la secuencia de decisiones acerca de cómo recoger, ordenar y analizar los datos en función de los fines de programa» (Anguera, 1992). De ahí que el término «diseño» se utilice como sinónimo de plan o estrategia metodológica que orienta el proceso evaluativo, lo cual implica realizar una planificación detallada sobre los aspectos o dimensiones del objeto a evaluar que vamos a considerar, los procedimientos que vamos a utilizar para recabar y analizar la información pertinente, y los sistemas a emplear para dar validez a las informaciones que van a sustentar los juicios de valor que se formulen.

Al igual que sucede con los modelos, también podemos utilizar diversos criterios a la hora de clasificar los diseños evaluativos (Ruttman 1984, Fitz-Gibbon y Morris 1987, Rossi y Freeman 1989, Scriven 1991). Entre estos criterios queremos destacar los tres que se utilizan con mas frecuencia:

7.2.1. En función de enfoque teórico adoptado

Utilizando este criterio los diseños se suelen clasificar de forma dicotómica en función de que el marco de la planificación se establezca tomando como referencia la evaluación de los efectos del programa o el significado que éste ha tenido para los implicados. Esta es la postura de adoptan algunos autores (Guba y Linclon 1981, Cronbach 1982, Patton 1987, Rebien 1997) al plantear dos tipos de diseños frontalmente opuestos —*preordenado y respondente*— como representativos de dos formas diferentes de abordar los procesos evaluativos relativos a los programas sociales. Una evaluación preordenada parte de un modelo teórico formalizado que se especifica mediante un

plan en el que se detallan todos los elementos del proceso: objetivos, audiencias, información necesaria, instrumentos, fuentes, técnicas de análisis, criterios de comparación, etc. Por el contrario, la evaluación respondente se caracteriza por partir de marco teórico flexible que busca las valoraciones de las personas implicadas en el programa para lo cual es necesario efectuar una planificación emergente que se pueda reconducir en cada etapa del proceso en función de las observaciones que se deduzcan de la etapa anterior.

7.2.2. El sistema de control que permitan las variables del programa

Desde un punto de vista metodológico estricto, el concepto de diseño alude a la «estrategia mediante la cual se pretende probar el efecto de una o más variables manipuladas»; es decir, un plan que nos permita constatar los efectos de las intervenciones promovidas mediante un programa. De ahí que cuando se utiliza el concepto de diseño con terminología positivista, su ordenación y clasificación se establece a partir de los tres requisitos o condiciones que debe reunir todo experimento, a saber: aleatorización de los sujetos, manipulación de la variable independiente y grupo de control. Si trasladamos estos requisitos a la metodología de evaluación de programas, los diseños evaluativos puede ser clasificados en cuatro grandes grupos en función del grado o nivel de garantía con que se cumplen estos requisitos. Así, de acuerdo con Anguera (1992, 1995), los estudios evaluativos se pueden clasificar en función de la metodología utilizada en los siguientes tipos: *observacionales, selectivos, cuasiexperimentales y experimentales*. Consideramos que a esta clasificación, aunque es muy comprensiva, se debería añadir la metodología centrada sobre *costos-beneficios*, cuyos diseños se utilizan cuando la finalidad del proceso se orienta a estimar la eficiencia del programa.

7.2.3. El sistema de control que se utiliza sobre la información recabada

Finalmente elegir un diseño implica determinar el procedimiento mediante el cual vamos a contrastar la información que recabamos sobre el programa —global o parcialmente— a fin de obtener las evidencias necesarias. Para ello no sólo cabe recurrir a los procedimientos propios de los estudios experimentales sino que también existen otros sistemas de control que —aunque no sean tan potentes— pueden ser más adecuados y oportunos para ser utilizados en determinadas fases y/o tipos de programas. En esta línea positivista, algunos autores (Rossi y Freeman, 1989) clasifican los diseños utilizando como criterios la existencia de grupo de control y el procedimiento utilizado para ello. Así hablan de diseños comparativos o de cobertura parcial mediante controles aleatorios, construidos y estadísticos, y diseños no comparativos o cobertura total que utilizan controles genéricos, reflexivos y en la sombra. Desde enfoques participativos los sistemas de control no se fundamentan tanto en la estructura del diseño como en las técnicas (juicios de expertos, comisiones ad hoc, triangulación, etc..) que se utilizan para contrastar y consensuar la información procedente de las diversos estudios y audiencias (Pineault y Davaluy 1989, Apodaca 1999).

307

7.3. Correspondencia entre modelos y diseños

Hemos avanzado que entre modelos y diseños existe una interacción evidente que se hace igualmente patente en el caso de la evaluación de un programa. Establecido el modelo, la elaboración del diseño constituye, en parte, una consecuencia lógica. Cuando la evaluación se enfoca desde una aproximación positivista lo lógico es que elaboremos un diseño preordenado y que utilicemos una metodología lo más coherente posible con los supuestos de este paradigma aplicando sistemas de control —a ser posible— comparativos. Por el contrario, si nuestro enfoque se establece desde posiciones más pluralistas lo normal es que la planificación del diseño se organice de forma más flexible fundamentalmente a partir de estudios observacionales y utilizando procedimientos de control participativos y no comparativos.

Así pues, dado el abanico de modelos y diseños posibles, el evaluador deberá concretar en cada caso la estrategia metodológica que considera más apropiada en función de su reflexión y análisis sobre los supuestos teóricos que justifican toda evaluación, pero teniendo en cuenta que entre modelo y diseño debe existir una clara coherencia, tal como hemos plasmado en otro trabajo anterior (De Miguel, 2000). Especialmente deberá valorar —a la hora de tomar una opción— las consecuencias prácticas, ya que la finalidad última en toda evaluación de programas es la toma de decisiones pertinentes que incidan sobre la mejora del programa. De ahí la necesidad de potenciar enfoques evaluativos integradores que tengan en cuenta tanto los efectos del programa como su significado para los implicados ya que ambos aspectos son esenciales a la hora de elaborar las propuestas de mejora. Ello implica, de nuestro punto de vista, que sea aconsejable abordar siempre la evaluación de un programa con un enfoque comprensivo que integre aproximaciones teóricas diversas y estrategias metodológicas complementarias. Lo cual significa optar por modelos y diseños menos radicales y más eclécticos.

8. CONSIDERACIONES FINALES

Una vez que hemos abordado «algunos» de los problemas que conciernen a la fundamentación teórica de la evaluación de programas y a las exigencias que conlleva la elección de un enfoque o aproximación metodológica, procede efectuar ciertas consideraciones finales que, a modo de resumen, contribuyan a clarificar el complejo marco en el que se mueve el desarrollo de esta disciplina y el trabajo de un evaluador. Como ya hemos dicho, la evaluación de programas es una tarea esencialmente práctica pero ello no significa que esté alejada de la teoría. Al contrario, necesitamos más y mejor teoría que ilumine lo qué hacemos, cómo lo hacemos y por qué. En definitiva, necesitamos justificar teóricamente una actividad que es esencialmente aplicada.

En este proceso de búsqueda hacia una fundamentación teórica, consideramos oportuno recapitular algunas de las lecciones que hemos aprendido desde la experiencia (Shadish, Cook y Leviton 1991, Chelimsky 1997, Cook 1977, Scriven 1997, Stake y Mabry 1997, Rossi, Freeman y Lipsey 1999, De Miguel 1999b) y presentar algunas de las connotaciones más relevantes que actualmente enmarcan la evaluación de toda

estrategia de intervención social. Señalamos, entre ellas, aquellas que nos parecen que pueden contribuir positivamente a consolidar la identidad científica de esta disciplina. A saber:

1. *Profundizar sobre las teorías relacionadas con el cambio social*

Abordar la evaluación de una estrategia de intervención social exige previamente precisar qué se entiende por cambio o mejora social y cómo debemos proceder para que esto sea posible. Ello requiere un conocimiento de las teorías sobre cómo se pueden resolver los problemas sociales y, en consecuencia, sobre las características que debe reunir toda estrategia de intervención que se oriente a promover determinadas mejoras. En esta línea algunos autores señalan la necesidad de que quienes planifican los programas formulen previamente una «teoría normativa» que defina y justifique cada una de las intervenciones y cómo se debe aplicar para que tenga las consecuencias deseadas (Wholey, Hatry y Newcomer, 1994). El evaluador que tiene un mayor conocimiento de las teorías sobre el cambio social estará en mejores condiciones de llevar a cabo procesos evaluativos más adecuados.

2. *Establecer una teoría evaluativa explícita además de implícita*

Toda evaluación requiere una reflexión teórica previa sobre el tipo de programa objeto de análisis con el fin de justificar los modelos y procedimientos que vamos a utilizar para llevar a cabo los procesos evaluativos. Generalmente esta reflexión no se lleva a cabo de forma explícita, aunque de una u otra forma todos los evaluadores tienen una teoría sobre la evaluación de forma implícita. Por ello, a efectos de que nuestra actividad sea algo más que una metodología aplicada, necesitamos reflexionar sobre los diversos supuestos que definen nuestro concepto sobre el programa y explicitar los parámetros teóricos y criteriales a partir de los cuales vamos a elegir el enfoque evaluativo. Necesitamos pasar de una teoría tácita o implícita sobre el programa a una teoría articulada que oriente y justifique la aproximación metodológica elegida (Weis 1997, Rossi, Freeman y Lipsey 1999).

3. *Clarificar los fundamentos epistemológicos de toda evaluación*

Ahora bien, para llegar a la formulación de una teoría explícita es necesario que, antes de proceder a la planificación de una evaluación, el evaluador reflexione y precise el punto de vista que adopta en relación con los diferentes supuestos o fundamentos que delimitan y justifican la evaluación de un programa concreto. De modo especial deberá concretar el marco de referencia que va a utilizar en el proceso de construcción del conocimiento; es decir, delimitar qué pretende conocer, qué tipo de conocimientos necesita y cómo va a proceder para su construcción. Esta concreción lógicamente implica reflexionar antes sobre cuestiones básicas de carácter ontológico, epistemológico y metodológico que son claves para poder entender y explicar sus decisiones prácticas (Shadish, Cook y Leviton 1991, Rebien 1997). Suplir esta reflexión por la «afiliación» a un paradigma no contribuye al desarrollo del pensamiento crítico del evaluador.

4. *Abordar la evaluación como una tarea ecléctica*

Este recurso —habitualmente utilizado— de definir o explicitar los supuestos teóricos dentro de un paradigma no es recomendable en el campo de la evaluación porque,

como ya hemos avanzado, no existe un paradigma que ofrezca un marco teórico comprensivo capaz de integrar todos los interrogantes y problemas que conlleva la realización de un proceso evaluativo que por su propia naturaleza debe responder a distintas finalidades, diferentes audiencias y usos muy diversos. La evaluación es esencialmente una tarea ecléctica por lo que es muy cuestionable la utilización de los paradigmas clásicos con este fin (Cook, 1997). Este hecho nos advierte, al mismo tiempo, en la necesidad de profundizar sobre el estatuto científico de esta disciplina desde otras perspectivas más realistas (Pawson y Tilley, 1997).

5. *Elegir los métodos en función de la finalidad*

Dado que no existe un prototipo de paradigma para evaluar programas, procede insistir que los métodos deben ser elegidos en función de la finalidad que orienta el proceso y no al revés como frecuentemente sucede. La experiencia nos lleva a afirmar que no existen métodos de evaluación que se puedan aplicar siempre y en todas las situaciones sino que éstos deben ser elegidos en función del propósito hacia la que se orienta la evaluación y la necesidad de obtener resultados «creibles». Así como los métodos cualitativos no son buenos para generar estimaciones cuantitativas sobre los resultados de un programa, tampoco los métodos cuantitativos son adecuados para obtener informaciones respecto a la teoría que lo fundamenta y los procesos desarrollados. No cabe, por tanto, apriorismos o planteamientos exclusivos ya que son necesarios ambos tipos de metodologías. Las dificultades se presentan en el plano de la logística, no en el de la lógica (Cook y Reichardt 1986, Cook 1997, Rossi, Freeman y Lipsey 1999).

6. *Fomentar la diversidad metodológica*

Cuando se asume que las evaluaciones deben responder a diversos propósitos y audiencias, lo habitual es utilizar una combinación parsimoniosa de datos, análisis y métodos en función del tipo de informaciones y conocimientos que perseguimos en cada momento y fase del proceso. La evaluación es un campo en el que se pueden aplicar todo tipo de estudios y metodologías y donde la tarea se enriquece precisamente con la pluralidad. Por ello, la única opción válida es abordar los procesos con un enfoque comprensivo y utilizar los procedimientos que en cada caso estimamos oportunos, al margen de polémicas estériles que no conducen a ninguna parte. La evaluación es un campo del conocimiento donde el debate entre metodologías está superado (Reichardt y Rallis 1994, Datta 1997, De Miguel 1999).

7. *Promover trabajos de síntesis antes que individuales*

Precisamente porque la mayoría de las evaluaciones exigen recabar informaciones diversas, utilizando fuentes muy distintas y métodos diferentes, parece necesario abordar esta tarea con una visión pluralista e integradora con el fin de que los juicios de valor se fundamenten sobre conocimientos contrastados por la opinión mayoritaria de todos los implicados en el programa. De ahí que la fiabilidad y validez de las observaciones y recomendaciones aumente en la medida que hayan sido elaboradas a partir de la integración de los resultados de varios estudios, lo que significa priorizar los trabajos de síntesis y metaanálisis frente a los individuales (Lipsey y Wilson 1993, Cook 1997). No debemos olvidar que la calidad de un conocimiento se incrementa en la medida que haya un escrutinio público mayor sobre el mismo.

310

8. *Velar por la credibilidad de los informes*

Fácilmente se comprende que la elaboración del informe sobre una evaluación no es una cuestión neutra en relación con las aplicaciones o usos de las recomendaciones sugeridas. Para que el informe o informes sean útiles deben reunir dos notas que nos parecen muy importantes. De una parte, deben responder a las necesidades de las audiencias y ser elaborados de forma objetiva que no ponga en cuestión la independencia del evaluador, lo que conlleva considerar la naturaleza de las relaciones entre el evaluador y los implicados. De otra, deben ser redactados en términos claros y con datos precisos que sean fácilmente comprensibles por las audiencias implicadas. Ambas cuestiones condicionan la credibilidad de los informes y, en consecuencia, su utilidad (Patton 1986, Wholey, Hatry y Newcomer, 1994).

9. *Analizar los factores contextuales y políticos que condicionan el uso de la información*

Además de la credibilidad que ofrecen los informes existen otros factores contextuales y políticos que condicionan la aplicación o utilización de los resultados y que también deben ser considerados por los evaluadores con el fin de eliminar las barreras que dificultan que los procesos tengan consecuencias prácticas. Entre estos factores debemos plantearnos todos los problemas que surgen entre la entrega de los informes y la puesta en práctica de las recomendaciones oportunas; es decir, la difusión y diseminación de los informes, la interpretación de resultados, el conocimiento de los implicados en la toma de decisiones, los argumentos que avalan cada una de las alternativas sugeridas, etc.. El análisis de estos problemas nos permitirá buscar soluciones orientadas a eliminar o reducir sus efectos e incrementar así la eficacia de nuestros trabajos (Braskamp y Brown 1980, Ewell 1989, Torres, Preskill y Piontek 1996).

10. *Realizar estudios de metaevaluación sobre los trabajos efectuados*

Finalmente todos debemos considerar que la mejor forma de avanzar y profundizar sobre el marco teórico de nuestra disciplina y nuestra tarea es a través de la propia evaluación del trabajo que realizamos. De ahí que debamos someter a control todos nuestros trabajos con el fin de que, al detectar sus puntos fuertes y débiles, podamos tener una visión realista sobre la calidad del mismo y las limitaciones que presenta para su utilización posterior. Para ello debemos habituarnos a efectuar estudios de metaevaluación utilizando guías elaboradas con este fin y remitir nuestros informes a evaluación externa para recabar otras opiniones independientes. Es un camino que necesariamente debemos recorrer si queremos mejorar nuestro trabajo, mejorar las bases teóricas de nuestra disciplina, y ganar más credibilidad ante nuestras audiencias (Joint Committee 1988, Shadish, Newman, Scheirer y Wye 1995).

Finalmente, quisiera añadir que estas consideraciones que sugerimos son algunas de las reflexiones a las que hemos llegado desde nuestra experiencia personal y que, en la medida que están en línea con el pensamiento de otros autores, esperamos puedan ser útiles para quienes están preocupados por estas cuestiones. Los evaluadores presionados por la práctica hemos dejado poco espacio para la reflexión teórica y quizás ha llegado el momento en que nos tengamos que plantear la necesidad de fundamentar nuestra tarea ya que en ello no sólo se haya comprometida la identidad de esta disciplina sino también la misma credibilidad de nuestro trabajo.

BIBLIOGRAFÍA

Alkin, M. y Ellett, J. (1990). «Development of Evaluation Models». En Walberg, H. y Haertel, G, (eds.). *The international encyclopedia of educational evaluation*. Oxford, Pergamon Press.

Anguera, T. (1992). «Diseños diacrónicos en programas de intervención». En *Rev. Bordón*, 43 (4), 421-429.

Anguera, T. (1995). «Diseños». En Fernández Ballesteros, R. (ed.). *Evaluación de programas*. Madrid, Síntesis.

Apodaca, P. (1999). «Evaluación de los resultados y del impacto». *Revista de Investigación Educativa*, 17(2), 363-377.

Bickman, L. (ed.) (1987). *Using Program Theory in Evaluation. New Directions for Program Evaluation nº 33*. San Franciso, Jossey-Bass.

Bickman, L. (ed.) (1990). *Advnances in Program Theory. New Directions for Program Evaluation nº 47*. San Franciso, Jossey-Bass.

Blaskar, R. (1975). *A realist theory of science*. Brighton, Hasvester.

Boruch, R. (1997). *Randomized controlled experiments for planning and evaluation: A practical guide*. Thousand Oaks, CA: Sage.

Braskamp. L. y Brown, R. (1980). *Utilization of Evaluation Information. New Directions for program Evaluation* nº 5. S. Francisco, Jossey-Bass.

Bryk, A. (ed.). (1983). *Stakeholder-based Evaluation. New Directions for program Evaluation* nº 17, S. Francisco, Jossey-Bass.

Chelemsky, E. (1997). «The Coming transformations in Evaluation». En Chelemsky, E. y Sadish, W. (eds.). *Evaluation for the 21 Century*. London, CA: Sage.

Chen, H.T. (1990). *Theory-driven evaluations*. Newbury, CA: Sage.

Cook, T. (1997). «Lessons learned in Evaluation Over the past 25 Years». En Chelemsky, E. y Sadish, W. (eds.). *Evaluation for the 21 Century*. London, CA: Sage.

Cook, T. y Reichardt, CH. (1986). *Métodos cualitativos y cuantitativos en la investigación evaluativa*. Madrid, Morata.

Cronbach, L. (1982). *Designing evaluations of educational and social programs*. S, Francisco, CA: Jossey-Bass.

Datta, L. (1997). «Multimethod Evaluations: Using case studies together with other methods». En Chelemsky, E. y Sadish, W. (eds.). *Evaluation for the 21 Century*. London, CA: Sage.

De Miguel, M. (1989). «Modelos de evaluación sobre organizaciones educativas». En *Revista de Investigación Educativa*, 13, 21-56.

De Miguel, M. (1990). «La investigación en la acción. ¿Una alternativa metodológica para las Ciencias Sociales? En Vázquez, J. y Otros. *Metodología de la Investigación científica*. Santiago de Compostela, ICE.

De Miguel, M. (1991). «Utilización de indicadores en la evaluación de la docencia». En De Miguel, M.; Mora, J. y Rodríguez, S. (eds.). *La evaluación de las Instituciones Universitarias*. Madrid, Consejo de Universidades.

De Miguel, M. (1995). «Participación ciudadana en la evaluación de las políticas públicas». *Aula Abierta*, 66, 211-218.

De Miguel, M. (1999). «La evaluación de programas: entre el conocimiento y el compromiso». *Rev. de Investigación Educativa*, 17 (2), 345-48.

De Miguel, M. (1999b). «Problemas teóricos y metodológicos en la evaluación externa de programas multiculturales». En Martin, T. y Serrate, L. (Coord.). (1999). *Evaluación y Ámbitos Emergentes en Animación Sociocultural*. Madrid, Sanz y Torres.

De Miguel, M. (2000). «Modelos y diseños en la evaluación de programas». En Sobrado, L. (ed.). *Orientación Profesional: Diagnóstico e Inserción Sociolaboral*. Barcelona, Editorial Estel.

De Miguel, M. (2000b). «La evaluación externa de un programa de Educación Social» En Pérez Serrano, G. (Coord.). *Modelos de Investigación Cualitativa*. Madrid, Narcea.

Dennis, M. y Boruch, R. (1989). «Randomized experiments for planning and testing projects in developing countries». *Evaluation Review*, 13 (3), 292-309.

Drummond, M., Stoddart, G. y Torrance, G. (1991). *Métodos de evaluación económica de los programas de atención a la salud*. Madrid, Diaz de Santos.

Ewell, T. (ed.). (1989). *Enhancing Information Use in Decision Making. New directiones for institutional research*, nº 64. S. Francisco, Jossey-Bass.

Fetterman, D., y Pitman, M. (1986), *Educational Evaluation*. London, Sage.

Fetterman, D. (1994). «Empowerment evaluation». *Evaluation Practice*, 15(1), 1-15.

Fetterman, D., Kaftarian, S. y Wasdersman, A. (eds.)(1966). *Empowerment evaluation*. London, CA: Sage.

Fitz-Gibbon, C. y Morris, L. (1987). *How to design a program evaluation*. Newbury Park, CA: Sage.

Folc'hlay, B. y Plottu, E. (1998). «Democratic evaluation». *Evaluation*, 4 (3), 261-277.

Guba, E. y Lincoln, Y. (1981). *Effective evaluation*. S. Francisco, Jossey-Bass.

Guba, E. y Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA.: Sage.

Herman, J., Morris, L. y Fitz-Gibbon, C. (1987). *Evaluator's handbook*. Newbury Park, CA: Sage.

Hesse, M. (1974). *The structure of scientific inference*. London, MacMillan.

House, E. (1980). *Evaluating with validity*. Newbury Park, CA.: Sage.

Howe, K. (1992). «Gedtting over the quantitative-qualitative debate». *American Journal of Education*, 100(2), 236-255.

Joint Committee On Standars, (1988). *Normas de Evaluación*. México, Trillas.

Keeves, J. (ed.). (1988). *Educational Research, Methodology and Measurement: A International Handbook*. Oxford, Pergamon Press.

Lewin, H.M. (1983). *Cost-effectivenes* Newbury Park, CA: Sage.

Lincoln, Y. y Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA.: Sage.

Lipsey, M. y Wilson, D. (1996). *Toolkit for Practical Meta-Analysis*. Cambridge, MA: Human Services Research Institute.

MacDonald, B. (1977). «A Political classification of evaluation studies». En Hamilton, D. y al. (eds.). *Beyond the Numbers Game*. London, MacMillan.

Madaus, G., Scriven, M. y Stufflebeam, D. (eds.) (1983). *Evaluations models: Viewpoints on educational and human services evaluation*. Boston, Mass.: Kluwer-Nijhoff.

Martin, L. (1993). *Total quality management in human service organizations*. Newbury Park, CA: Sage.

Martin, L. y Kettner, P. (1996). *Measuring the performance of human service program*. London, Sage/SHSG.

McLaughlin, G. y McLaughlin, J. (1989). «Barriers to Information Use: The Organizational Contex». En Ewell, T. (ed.). *Enhancing Information Use in Decision Making. New directiones for institutional research*, nº 64. S. Francisco, Jossey-Bass.

McLaughlin, M y Phillips, C. (eds.) (1991). *Evaluation and Education: At quarter Century*. Chicago, NSSE.

Nas, T. (1996). *Cost-benefit analysis. Theory and application*. Newbury Park, CA: Sage.

Nevo, D. (1990). «Role of the evaluator». En Walberg, H. y Haertel, G. (eds.): *The international encyclopedia of educational evaluation*. Oxford, Pergamon Press.

Patton, M. (1986). *Utilization-focused evaluation*. (1ª ed., 1978). Newbury Park, CA: Sage.

Patton, M. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA.: Sage.

Patton, M. (1994). «Developmental Evaluation». *Evaluation Practice*, 15(3), 311-319.

Pawson, R. y Tilley, N. (1997). *Realistic Evaluation*. Newbury Park, CA.: Sage.

Pineault, R. y Daveluy, C. (1989). *La planificación sanitaria: conceptos, métodos y estrategias*. Barcelona, Masson.

Rebien, C. (1997). «Development assistance evaluation and the foundations of program evaluation». *Evaluation Review*, 21 (4), 438-460.

Reichart, C. y Rallis, S. (eds.) (1994). *The qualitative-quantitative debate: new perspectives. New directions for Evaluation*, nº 61. S. Francisco, Jossey Bass.

Rossi, P. y Freeman, H. (1989). *Evaluación. Un enfoque sistémico para programas sociales*. México, Trillas.

Rossi, P., Freeman, H. y Lipsey, M. (1999). *Evaluation. A systematic approach*. Thousand Oaks, Sage.

Ruttman, L. (1984). *Evaluation research methods: a base guide*. Beverly Hills, Sage.

Sayer, A. (1984). *Method in social science*. London, Hutchinson.

Schalock, R. (1995). *Outcomes-based evaluation*. Newbury Park, CA.: Sage.

Scriven, M. (1991). *Evaluation Thesaurus*. Newbury Park, CA.: Sage.

Scriven, M. (1997). «Truth and Objectivity in Evaluation». En Chelimsky, E. y Sadish, W. (eds.). *Evaluation for the 21 Century*. London, CA: Sage

Shadish, W., Cook, T. y Leviton, L. (1991). *Foundations of program evaluation*. Newbury Park, CA.: Sage.

Shadish, W., Newman, D., Scheirer, M. y Wye, C. (eds). (1995). *Guiding Principles for Evaluators. New Directions for Evaluation*, nº 66. S. Francisco, Jossey-Bass.

Sheerens, J. (1987). «Beyond decision-oriented evaluation». *International Journal of Educational Research*, 11 (1) 83-93.

Smith, M. (1998). «Empowerment evaluation: Theorical and methodological considerations» *Evaluation and Programming Planning*, 21, 255-261.

Sonnichsen, R. (1994). «Evaluators as Change Agents». En Wholey, J., Hatry, H. y Newcomer, K. (1994). *Handbook of Practical Program Evaluation*. S. Francisco, Jossey-Bass Publishers.

Stake, R. y Mabry, L. (1997). *Advances in program evaluation*. Newbury Park, CA.: Sage.

Stecher, B. y Davis, W. (1991). *How to focus an evaluation*. Newbury Park, CA.: Sage.

Stufflebeam, D. y al. (1971). *Educational Evaluation and decision-making in Education*. Illinois, Peacock-Itasca.

Stufflebeam, D. y Shinkfield, A. (1987). *Evaluación sistemática*. Barcelona, Paidós/MEC.

Stufflebeam, D. (1994). «Empowerment evaluation, Objectivist evaluation and Evaluation Standars». *Evaluation Practice*, 15, 321-338.

Torres, R., Preskil, H. y Piontek, M. (1996). *Evaluation Strategies for Communicating and Reporting*. Thousand Oaks, CA: Sage.

Vedung, E. (1997). *Evaluación de políticas públicas y programas*. Madrid, Instituto Nacional de Servicios Sociales.

Walker, J. y Evers, C. (1988). «The epistemological unity of educational research». En Keeves, J. (ed.). *Educational Research, Methodology and Measurement: A International Handbook*. Oxford, Pergamon Press.

Weis, C. (1975). *Investigación evaluativa*. México, Trillas.

Weis, C. (1997). «How can Theory-based Evaluation make Greater Headway?». *Evaluation Review*, 21(4), 501-524.

Wholey, J., Hatry, H. y Newcomer, K. (1994). *Handbook of Practical Program Evaluation*. S. Francisco, Jossey-Bass Publishers.

Worthen, B. y Sanders, J. (1987). *Educational evaluation*. London, Longman.

Worthen, B., Sanders, J. y Fiztpatrick, J. (1997). *Program evaluation alternative approaches and practical guidelines*. Newbury Park, CA.: Sage.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 319-339

EL DISEÑO Y LOS DISEÑOS EN LA EVALUACIÓN DE PROGRAMAS

Fco. Javier Tejedor
Universidad de Salamanca

1. LA EVALUACIÓN DE PROGRAMAS COMO INVESTIGACIÓN EVALUATIVA

La Evaluación de Programas es un proceso para generar formas útiles de comprensión sobre una innovación (o una intervención). Este sentido de utilidad hay que entenderlo como la posibilidad de utilizar la intervención como un recurso crítico para mejorar los procesos de acción educativa, razón por la que, a nivel metodológico, puede entenderse como un proceso de investigación evaluativa (Tejedor y otros, 1994; Pérez Juste, 1995; Amezcua y Jiménez, 1996; Martínez Mediano, 1997).

La investigación evaluativa ha ido perfilándose con una metodología concreta en los últimos quince años. Esta metodología queda delimitada por: una terminología propia, un conjunto de herramientas conceptuales y analíticas específicas, y unos procesos y fases también específicos. El conjunto define un campo floreciente en el que realmente lo específico es la mezcla y no las partes constitutivas. Para Alvira (1991), algunas circunstancias que han contribuido a conformar este campo de actuación serían:

a) Mayor interrelación entre las actividades de evaluación y programación. Si en una perspectiva clásica la evaluación se realiza después de las actividades de planificación, programación y desarrollo de la intervención, en la actualidad ambas actividades caminan de forma paralela.
b) La investigación evaluativa es investigación (con todos los considerandos epistemológicos, metodológicos y técnicos que ello entraña), pero es más que investigación, ya que se entiende como un proceso cuya finalidad es aplicar procedimientos científicos para acumular evidencia válida y fiable sobre la manera y grado en que un conjunto de actividades producen resultados o efec-

tos concretos. Intenta determinar el valor de la acción educativa, buscar soluciones a problemas concretos o mejorar las estrategias de actuación.

c) Se produce una mayor implicación de las partes interesadas (observación participante, investigación acción,...).
d) Pueden integrarse estrategias de investigación muy diversas, tanto cuantitativas como cualitativas. Es un campo de trabajo que posibilita la utilización simultánea y sucesiva de diferentes metodologías para el tratamiento de los temas objeto de estudio e intervención. Será la naturaleza específica del programa y la pericia del evaluador quienes a la postre determinarán la utilización de una o varias opciones.

La investigación evaluativa es un modo de investigación que implica un proceso riguroso, controlado y sistemático de recogida y análisis de información fiable y válida para tomar decisiones sobre un programa educativo. Vemos que aparecen importantes puntos de contacto respecto a lo que entendemos por «investigación»: proceso sistemático, riguroso, controlado; utilización de técnicas de medición y análisis de datos. La evaluación se distingue de la investigación por los objetivos y no por los métodos...

Algunos autores, sin embargo, creen que conviene diferenciar ambas tareas, señalándose como características peculiares de la investigación evaluativa las siguientes (De la Orden, 1985):

a) Los juicios de valor acompañan al desarrollo de todo el programa: desde la selección del problema a estudio hasta en la aplicación de metodologías concretas.
b) Resulta difícil (a veces inadecuado) formular hipótesis precisas.
c) La replicación es prácticamente imposible dadas las características casi irrepetibles de la ejecución de un programa de intervención.
d) La recogida de datos está condicionada por la viabilidad del proceso.
e) Multitud de variables relevantes sólo son susceptibles de un control superficial limitándose por tanto la utilización de diseños de carácter mas experimental.
f) Las decisiones importantes en relación con la paralización, sustitución o repetición del programa no son competencia del evaluador.
g) El informe debe adaptarse a las exigencias de quien lo financia.

Una característica básica de la investigación evaluativa es que se lleva a cabo en un proceso de intervención, en la misma acción, que es lo realmente importante. La investigación, señalaba Weis (1982), es una prioridad secundaria, lo que debe llevarnos a pensar que la evaluación/investigación tiene que adaptarse al contexto del programa y perturbar lo menos posible su desarrollo.

El proceso de evaluación de programas debe concebirse como una auténtica estrategia de investigación sobre los procesos educativos, en cuyos resultados deberían basarse las pautas sugeridas para orientar los procesos de intervención. Investigación e intervención se apoyan mutuamente. La intervención sacará partido de la investigación en la medida en que ésta se extienda a la solución de problemas prácti-

317

cos. La investigación activa se asimila, enriqueciéndola, a la noción de investigación aplicada.

Toda situación educativa es susceptible de este doble tratamiento. Por una parte, existe la necesidad de «conocerla» y de «explicarla». Por otra, es preciso «comprenderla» y «mejorarla». Se requiere que la investigación alcance un nivel de integración suficiente, en este caso, entre el pensamiento y la decisión: la preocupación por el saber se une de alguna manera a la preocupación por el actuar. Este campo de estudio, que algunos denominan «investigación orientada», está a la base de toda investigación evaluativa.

La investigación orientada/evaluativa nace como respuesta a necesidades sociales y se elabora en función de tales necesidades; se realiza con objeto de llegar a la solución de los problemas que exige una acción informada científicamente. Habrá de tener necesariamente un carácter interdisciplinar y viene determinada por los objetivos, los resultados y las motivaciones.

Al referirnos al tema del diseño en la Evaluación de Programas nos vamos a encontrar, como ocurre en el más amplio campo de la investigación, con una doble acepción de dicho término:

a) Diseño como planificación general del proceso de intervención (el diseño o diseño de evaluación)
b) Diseño como planificación puntual del análisis de los efectos de la intervención realizada (los diseños o diseños de investigación).

Esta diferenciación es la que se ha tenido en cuenta a la hora de titular esta aportación y que pasamos a comentar con detenimiento.

2. EL DISEÑO EN LA EVALUACIÓN DE PROGRAMAS

Un diseño es un plan que fija qué medidas y cuándo serán tomadas en el curso de un proceso de evaluación. Plantearse la especificación de un diseño supone haber realizado reflexiones previas en torno a cuestiones relacionadas con el concepto general de evaluación y con las finalidades de las diferentes opciones metodológicas; posteriormente delimitaremos, como parte de la planificación prevista, los fundamentos de la misma, los objetivos, los elementos personales implicados en el programa, los ámbitos de intervención, los instrumentos de recogida de datos, las técnicas de análisis y los momentos de la evaluación; todo ello, de acuerdo al modelo que se haya elegido para realizar la actividad evaluativa y teniendo en cuenta que en todos estos elementos existen diferencias según la perspectiva o enfoque (paradigma o modelo) por se siga.

El diseño de Evaluación de Programas es tanto un proceso como un conjunto de escritos o planes que incluyen unos elementos que presentan los propios contenidos y que generan numerosas preguntas. No hay un único camino para realizar la evaluación de un programa por lo que habrá necesariamente que prestar atención a numerosos elementos a la hora de concretar un diseño o plan de análisis: condiciones

318

organizacionales, necesidades, intereses y valores de los implicados en el proceso, objetivos del programa, recursos necesarios y disponibles,...

En cualquier caso, independientemente del modelo que se haya elegido, cualquier mínima planificación tiene que atender las siguientes tareas:

— establecer las actividades que se van a evaluar
— fijar los criterios de evaluación
— elegir las estrategias para la obtención de información
— analizar la información
— tomar decisiones a partir de la información analizada

Donde podemos identificar los dos elementos básicos del diseño de una evaluación, que expresados en términos sintéticos serían:

— La determinación de los datos que se requieren para valorar la eficacia del programa
— La determinación del plan para obtener la información necesaria que nos permita establecer relaciones entre el programa y los resultados obtenidos

En el cuadro 1 presentamos una reformulación simplificada de la propuesta de Pérez Carbonell (1998, pp. 188-192) respecto a los aspectos a considerar en la delimitación del diseño en el marco de la Evaluación de Programas.

Algunas cuestiones importantes a tener en cuenta serían (Carballo, 1990, págs. 166):

— ¿Se van a establecer las pautas que se seguirán en el proceso evaluativo? (diseños prefijados vs emergentes)
— Los resultados de la evaluación, ¿se van a utilizar para mejorar el programa, para emitir un informe sobre su funcionamiento o para ambas cosas a la vez? (diseños para una evaluación formativa vs normativa?
— ¿Se van a incluir en el estudio procesos de intervención o sólo se va a describir el funcionamiento del programa? Si se incluye un proceso específico de intervención será necesario analizar sus efectos, lo que requerirá la explicitación de un «diseño de investigación»

a) Diseños prefijados vs emergentes

Los diseños prefijados siguen la pauta del modelo positivista, requiriéndose que los objetivos de la evaluación estén claramente definidos y fijados con anterioridad a la ejecución del programa, lo que supone que todos los componentes del modelo (objetivos variables, tratamientos, fuentes de información, datos a recoger, análisis de datos,...) estén determinados a partir de presupuestos teóricos o de experiencias en evaluación de otros programas similares.

Estos diseños tratan de informarnos sobre la consecución de objetivos en la ejecución del programa, intentando el evaluador esclarecer los factores (variables y relacio-

CUADRO 1
ASPECTOS A CONSIDERAR EN LA DELIMITACIÓN DEL DISEÑO

Elementos	Contenidos	Dificultades	Consecuencias
Desarrollo del programa	Entorno Objeto de la evaluación Destinatarios Criterios de evaluación Actividades	No se considera a los implicados	No se implican los interesados y se tiende a descalificar la evaluación
Recogida de datos	Variables de interés Fuentes de información Instrumentos Técnicas de recogida	Procedimientos que violan derechos	Pleitos
		Procedimientos que utilizan tiempo valioso de usuarios	Falta de cooperación Datos no reales
		Procedimientos que implican cambios de hábitos	
Análisis de datos	Tipo de datos recogidos Criterios estadísticos Análisis adecuados Interpretación correcta	Datos con incidencia individual	Información criticada y no valorada
		Análisis demasiado técnicos	
Realización de informes	Destinatarios Número de informes Difusión	Los implicados no participan en el informe	La información no se utiliza por los grupos afectados
		La audiencia no tiene oportunidad de reaccionar	
		El informe afecta a puestos de trabajo	
		Se pone en aprieto a personas o instituciones	
Manejo de la información	Decisiones a tomar Personas afectadas Presupuesto de las decisiones a adoptar	No se especifican bien responsabilidades	Agravio colectivo
		Presupuestos deficientes Información desaprovechada	

nes entre ellas) que han mediado en dicha consecución. En síntesis, estos diseños «fijos» se caracterizan porque:

— son planeados con anterioridad
— están estructurados, aunque admite reajustes
— se especifican problemas y objetivos, fuentes de información, instrumentos, temporalización,...
— se fijan los destinatarios de los informes
— los datos suelen ser de carácter cuantitativo y se valora todo aquello que se aproxima al paradigma racionalista

La práctica evaluativa ha ido poniendo de manifiesto que la evaluación de la consecución de los objetivos, planteada en estos términos, resultaba muy difícil de realizar

BEST COPY AVAILABLE

ya que las cosas no suelen desarrollarse en los términos inicialmente previstos, lo que llevó a los evaluadores a flexibilizar sus planteamientos evaluativos iniciales, abriéndose a nuevas perspectivas valorativas, no sólo de los resultados sino también del propio proceso seguido (evaluación del proceso) intentando relacionar la valoración de los resultados con las condiciones y circunstancias en las que se desarrolló el programa, lo que nos conduce a criterios de eficiencia y no sólo de eficacia.

En seguida, este tipo de alternativa tuvo seguidores fervorosos (Stake, Owens, Parlett y Hamilton, Guba...) y, especialmente, Scriven que incorpora como novedad la idea de valorar los efectos no previstos, dejándose llevar por el desarrollo del programa y cuestionándose incluso la utilidad de la formulación inicial de objetivos (Scriven, 1973).

Estos diseños «emergentes» pueden, globalmente, caracterizarse en los siguientes términos:

— se adaptan mejor a las necesidades e intereses de los participantes
— están abiertos a posibles modificaciones que durante el desarrollo puedan incorporarse
— los problemas y objetivos no tienen por qué ser especificados inicialmente; se produce una reformulación constante de los objetivos del programa
— no se especifican elementos tales como instrumentos de recogida de datos, momentos de obtención de la misma,..
— se exige que el evaluador siga el desarrollo del programa estando en contacto permanente con los participantes; este distinto grado de implicación del evaluador va a condicionar las técnicas e instrumentos de obtención de datos
— los datos suelen ser de carácter cualitativo y se valora todo aquello que se aproxima a los paradigmas interpretativo y crítico
— debe cuidarse la adecuada selección de información (demasiado abundante e irrelevante en muchos casos) y una implicación excesiva del evaluador en la valoración de logros

Quizás puede pensarse que esta diferenciación dicotómica establecida en los diseños («fijo/emergente») es más teórica que práctica. En realidad parece posible pensar que no exista tal dicotomía; pensamos que el diseño, en el marco de la Evaluación de Programas, debe ser establecido a priori, respondiendo a la concepción teórica que sobre la evaluación tienen los responsables del programa y el propio evaluador, debiendo incluir referencias a los objetivos del programa, sus componentes, los recursos necesarios, las estrategias de recogida y análisis de datos, las posibles decisiones que de su desarrollo puedan derivarse y la presentación y difusión de los resultados.

Pero abogar por la previsión no es estar a favor de la inmutabilidad. Nuestra postura en modo alguno excluye la posibilidad de ir incorporando durante el desarrollo de la ejecución del programa todos aquellos elementos que se consideren más favorables a la optimización del proceso planificado: desde la reformulación de objetivos a la difusión de resultados pasando por la incorporación de nuevos recursos o la modificación de las estrategias de recogida y análisis de datos.

b) Diseños de evaluaciones formativas vs sumativas

Puede afirmarse que fue Cronbach quién, a partir de 1980, extendió la aplicación de los términos «formativa» y «sumativa», inicialmente propuestos por Scriven para su aplicación a contextos instructivos, al ámbito de la evaluación de programas. En este contexto, la evaluación formativa es una parte importante del proceso que sigue el desarrollo del programa, proporcionándole retroalimentación de su funcionamiento para favorecer el mismo e incorporar los cambios que puedan contribuir a mejorar su realización, lo que implica la vinculación del evaluador con el discurrir del programa: evaluación de carácter interno con efectos potenciales, sobre todo en la mejora del desarrollo del programa. Sin duda, responde al papel más noble que puede desempeñar la evaluación: su incardinación en el desarrollo del propio programa, de la propia intervención, del propio proceso educativo.

La evaluación formativa permite establecer una interacción continua entre la evaluación y la acción educativa, destacando su repercusión en la mejora del proceso y en la orientación del programa y, por consecuencia, en la optimización de los productos (logros) que se obtengan tras la aplicación de dicho programa, sobre todo si la evaluación formativa se plantea la búsqueda de las causas que están explicando los datos que nos vamos encontrando.

En síntesis, el diseño formativo puede caracterizarse por los siguientes considerandos:

— la evaluación se plantea desde una perspectiva interna
— su objetivo es intentar la mejora del programa
— los datos son utilizados por el propio equipo técnico del programa o por los implicados (el evaluador debe ser parte integrante del equipo)
— se persigue sobre todo la utilidad para los usuarios
— el diseño puede ser tanto fijo como emergente (si es fijo deberán aceptarse cambios)

A la hora de valorar un diseño de corte formativo deberá considerarse (Bartolomé, 1990):

— la coherencia lógica del diseño
— su capacidad para responder a las necesidades del grupo y/o a su proyecto social y educativo
— su flexibilidad de adaptación y mejora
— su posibilidad de realización, a través de un acercamiento progresivo y de las percibidas por el grupo y de un acercamiento de las actividades previstas y realizadas
— su eficacia progresiva, valorando su incidencia real en la mejora educativa

La evaluación sumativa, por su parte, se relaciona con la terminación del proceso de intervención, intentando globalizar la evaluación formativa que se haya podido ir

322

haciendo a lo largo del proceso. La evaluación sumativa nos informa sobre los resultados obtenidos y, por tanto, sobre las necesidades que han podido satisfacerse con la aplicación del programa. Parece importante destacar el hecho de relacionar la consecución de logros con la satisfacción de necesidades en mayor medida que con la consecución de objetivos, lo que puede ayudarnos a reconocer la especificidad de los procesos de intervención en contextos de carácter no estrictamente instructivo.

El diseño sumativo respondería a las pautas siguientes:

— la evaluación puede plantearse desde una perspectiva externa
— su objetivo es intentar ofrecer un juicio sobre la satisfacción de necesidades y/o sobre la consecución de los objetivos
— interesa en gran medida la credibilidad externa
— el evaluador debe ser ajeno al equipo técnico del programa para favorecer la objetividad en la valoración de los resultados
— el diseño suele ser mas bien de tipo fijo que emergente

c) Distintos tipos de evaluaciones

Desde una perspectiva metodológica, la puesta en marcha de un programa supone una actuación que debe atender ordenadamente a las distintas fases del proceso. La supervisión del desarrollo de cada una de esas fases en términos evaluativos nos lleva al establecimiento de diferentes tipos de evaluación que pueden llegar a tener interés en si mismo, incluso con su metodología propia (necesidades, objetivos, proceso, resultados,...); pero no podemos olvidar que todos ellos forman parte de un esquema único de investigación que nos informará sobre la puesta en marcha del programa y la calidad de sus logros. Este esquema global del proceso evaluativo de un programa de intervención sería similar al que presentamos a continuación.

1) Evaluación de necesidades

Debe ser la primera actividad a llevar a cabo en la puesta en marcha de un programa de intervención (Tejedor, 1990). Aunque algunos autores plantean la evaluación de necesidades con el mismo énfasis que la evaluación de programas, considero que es mas pertinente concebirla como una fase de dicho proceso, limitando entonces sus objetivos, lo que no implica desvirtuar su esquema operativo que entiendo está suficientemente conformado. Podemos pensar que la evaluación de necesidades, siendo una fase del proceso de investigación para la puesta en marcha de un programa de intervención, tiene entidad propia y es susceptible de ser analizada con un esquema completo de investigación evaluativa (Witkin, 1984).

2) Evaluación del diseño (evaluación de entrada)

Supone la preocupación por la secuencia de los pasos a seguir, por la claridad de formulación de los objetivos planteados, por la calidad de los instrumentos o técni-

cas de obtención de información, por la suficiencia de la muestra fijada, por la adecuada localización de las unidades muestrales, por la procedencia de las actividades sugeridas, por la adecuación de las técnicas de análisis,... (Mayer, 1985; Bartolomé, 1990). Se pretende valorar el plan de acción (diseño) y los recursos del programa, a fin de seleccionar las estrategias que permitan satisfacer más adecuadamente las necesidades detectadas y que mejor se adapten al grupo de personas al que el programa va dirigido.

Incluye igualmente la evaluación de las posibilidades de evaluación del propio diseño (evaluación de la evaluabilidad). La experiencia determina que pocos de los programas llevados a la práctica son evaluables al tener problemas de especificación de objetivos, de implementación, de adecuación de las actividades a la consecución de objetivos, de carencia de medios para su desarrollo,...Por ello estamos de acuerdo con quienes sugieren la conveniencia de llevar a cabo una revisión previa de las posibilidades evaluativas del programa (Anguera, 1990).

3) Evaluación de la implementación

Este tipo de evaluación pretende analizar el funcionamiento del programa y determinar si existen diferencias entre el funcionamiento real y el previsto inicialmente. Si hubiera diferencias notorias entre ambos esquemas de funcionamiento trataríamos de hallar las causas y, si fuera posible, subsanar las deficiencias (Finnegan y otros, 1988). Puede pensarse que incluye:

— Evaluación de la cobertura: Trataremos de analizar si el programa cubre todos los objetivos previstos y si llega a todos los sujetos para los que se había pensado, determinando, si procede, los sesgos de cobertura existentes y las causas que los han motivado.
— Evaluación del proceso (seguimiento): En algunos casos, sobre todo si el programa es de larga duración, conviene realizar un análisis de su funcionamiento a fin de verificar si su desarrollo sigue las pautas previstas. Implica la revisión de todos los elementos constitutivos del programa (personales, funcionales y materiales), en su dimensión interactiva: calidad de las actividades, distribución del tiempo y del trabajo, utilización de recursos,...

La evaluación adecuada del proceso supone la consideración de todos sus elementos por lo que se sugiere un esfuerzo previo de reconsideración respecto a las fases anteriores de ejecución del programa (necesidades, objetivos, diseño,...). Los elementos a considerar serán:

— utilización adecuada de recursos
— temporalización de actividades
— implicación de los destinatarios en las actividades
— obtención de información

4) Evaluación de resultados

Hasta hace poco tiempo era el único tipo de evaluación que se llevaba a cabo tras la finalización de las actividades que implicaba la ejecución del programa. En términos generales, se trata de analizar el grado en que el programa consigue los resultados previstos (también pueden evaluarse, siempre que se hayan previamente identificado, resultados no previstos inicialmente). Se intenta determinar hasta que punto pueden vincularse los logros obtenidos con los efectos del programa, lo que tiene especial interés en los diseños de intervención dada la dificultad para establecer nexos de causalidad y covariación entre las variables consideradas en el programa.

La evaluación de resultados admite la modalidad de «evaluación de verificación», que pretende efectuar comprobaciones a base de patrones claramente identificados, tanto en base a datos cuantitativos como cualitativos, y la modalidad de «evaluación comprensiva», que pretende ampliar o profundizar la comprensión de situaciones y que utiliza básicamente datos cualitativos.

La temática de la evaluación de resultados es amplia y compleja, apareciendo tratada en la bibliografía con diferentes terminologías: evaluación de productos, de la efectividad, de la eficacia, de la eficiencia,... (De la Orden, 1985, 1990). Los elementos a considerar en la evaluación de productos serán:

— calidad de la información obtenida
— posibilidad de valorar la consecución de objetivos a partir de la información disponible
— calidad de los análisis y adecuación a los datos disponibles
— propuestas para la toma de decisiones
— pertinencia de las decisiones tomadas

5) Metaevaluación

Los elementos a considerar serán las pautas de cambio estratégico en la intervención y en la planificación evaluativa

3. LOS DISEÑOS DE INVESTIGACIÓN EN LA EVALUACIÓN DE PROGRAMAS

La evaluación de efectos supone la elección de un diseño adecuado de investigación que permita obtener, analizar e interpretar la información relevante para formular un juicio de valor y toma de decisiones apropiadas. El diseño determina la posibilidad de detectar y comprender los procesos y el impacto de un programa educativo, dada la multiplicidad de factores que influencian su operación y resultados (De la Orden, 1990).

Los diseños de evaluación de programas pueden variar desde el más riguroso experimento de campo a la descripción naturalista, pasando por la diversidad de diseños cuasiexperimentales o correlacionales. Existen muchas propuestas de clasificación

de diseños de investigación utilizables en el ámbito específico de la Evaluación de Programas (Fitz Gibbon y Morris, 1987; Colás y Rebollo, 1993). Nuestra propuesta es la siguiente:

A) Diseños para la investigación cuantitativa

1) Diseños descriptivos-comparativos (estudios tipo survey)
2) Diseños longitudinales
   — secuenciales (diacrónicos)
   — transversales (sincrónicos)
3) Diseños Correlacionales:
   — correlación y regresión
   — técnicas de clasificación
4) Diseños causales
5) Diseños experimentales:
   — preexperimentales
   — cuasiexperimentales
   — experimentales de campo (intergrupo e intragrupo)

B) Diseños para la investigación cualitativa

1) Observación naturalista:
   — Estudios de casos
   — Estudios del grupo
   — Estudios de comunidades (etnográficos)
   — Estudios de situaciones o contextos

2) Diseños participativo-críticos
   — Investigación-acción

A) El entorno cuantitativo

El diseño experimental

Presenta las características propias del método científico (repetibilidad, estabilidad y comunicabilidad), siendo sus rasgos distintivos la manipulación de variables por parte del investigador y el control de variables intervinientes en el fenómeno estudiado.

Pretende llegar a establecer relaciones causales entre una o más variables independientes y una o más variables dependientes, lo que únicamente será posible a partir de un riguroso control de las circunstancias en que podemos llegar a establecer con claridad dicha relación causal.

Se estudian las modificaciones que sufren las variables dependientes a partir de cambios intencionados (manipulación o intervención) en las variables independientes.

La exactitud de la relación establecida entre ambas dependerá de la calidad del control ejercido sobre otras variables intervinientes, bien a partir de medios físicos (similitud de situaciones), bien a partir de técnicas aleatorias.

Los diseños que se enmarcan en el método experimental se caracterizan por tener una gran validez interna, que no es más que la capacidad para atribuir ciertamente la variación observada en la variable dependiente a la modificación intencional llevada a cabo sobre la variable independiente en base al control ejercido sobre el resto de las variables (Tejedor, 1994).

El grado de seguridad de los resultados de una investigación no debe juzgarse exclusivamente en términos de validez interna. Conviene reconsiderar también sobre otros tipos de validez: externa, estadística, de constructo y ecológica.

Los presupuestos exigidos a toda investigación en términos de valor de verdad, aplicabilidad, consistencia y replicabilidad, se reinterpretan en el método experimental en términos de validez interna, validez externa, fiabilidad y objetividad (en las estrategias de investigación activa y participativa se reinterpretarán en términos de credibilidad, transferibilidad, dependencia y conformabilidad).

Un aspecto importante a tener en cuenta por el evaluador que decida seguir algunas de las variantes del método experimental es el problema que representa la «experimentación con sujetos humanos». Desde los distintos efectos perturbadores que pueden modificar las conclusiones obtenidas hasta las interacciones que pueden producirse en las características personales de los sujetos. Pese a todo, las críticas más fuertes al método experimental le llegan por sus características de artificialidad, de reduccionismo de la realidad que pretende explicar y por la operacionalidad implícita de las variables que se incluyen. Por ello no consideramos adecuada la aplicación de los diseños experimentales, en sentido estricto, en el marco de la Evaluación de Programas.

Experimentos de campo

Sin embargo, sí creemos posible, aunque con muchas limitaciones, pensar en la modalidad de «experimentos de campo», que presentando los mismos presupuestos formales, su viabilidad de aplicación reside en el hecho de llevarse a cabo en marcos naturales, lo que supone una evidente reducción de su validez interna. Kerlinger (1977) caracterizaba los experimentos de campo en los siguientes términos:

— se realizan en condiciones naturales, interviniendo todas las variables que influyen en el fenómeno
— los efectos producidos por las variables son generalmente más fuertes que en el laboratorio
— existe mayor validez externa, por su referencia a situaciones más reales
— existe la posibilidad de estudiar fenómenos complejos en situaciones similares a la vida misma; son aplicables a gran variedad de problemas
— son apropiados para la demostración de teorías, comprobación de hipótesis y la resolución de problemas prácticos.

Entre las dificultades que puede encontrarse el investigador social en la aplicación de estos diseños cabría citar los aspectos relacionados con el control de variables, con la aleatorización de las muestras y con la disponibilidad de experimentadores bien entrenados.

Aunque son muchos los diseños que pueden aplicarse en los experimentos de campo, en todos ellos deben seguirse en su planificación los siguientes pasos:

— determinación de las muestras a utilizar;
— especificación de los tratamientos de cada una de las variables independientes;
— asignación de las condiciones experimentales (tratamientos) a los diferentes grupos;
— selección de las técnicas adecuadas de control de variables;
— constitución de los «grupos de control», si procede.

Cuasi-experimentos

Cuándo no es posible establecer un control efectivo sobre las variables intervinientes, el investigador social puede recurrir a diseños de control parcial o cuasi-experimentos. Deberá prestarse en su aplicación especial interés a las variables que no se han podido controlar y considerar la posibilidad de que los resultados que obtenga (relaciones entre las variables independientes y dependientes) puedan deberse más a la influencia de las variables no controladas que a los efectos de las variables independientes.

La carencia de control se produce básicamente por la imposibilidad de manipular la variable independiente y por las dificultades de asignar aleatoriamente los sujetos a las condiciones experimentales. Las principales características de un diseño cuasi-experimental serían:

— el empleo de escenarios naturales, generalmente de tipo social;
— la carencia de un control experimental completo;
— el uso de procedimientos (producción de observaciones múltiples) como sustitutos del control experimental intentando minimizar y, en algunos casos eliminar, los efectos de tantas fuentes de invalidez interna como sea posible;
— su disponibilidad, ya que pueden realizarse cuando no es posible la aplicación de un diseño experimental (de laboratorio o de campo).

Pueden considerarse aportaciones ingeniosas de los diseños experimentales a investigaciones naturales, que utilizan para la evaluación del cambio producido por las condiciones estudiadas grupos de control no equivalentes.

Los numerosos diseños cuasi-experimentales se pueden encuadrar en una de las dos modalidades siguientes:

1. *Diseños simples de series temporales.* Realizan medidas de la variable dependiente antes y después de la aplicación de la variable independiente; el posible efecto

de ésta puede ser estimado en función de la discontinuidad que presentan las medidas tomadas después, que se puede manifestar bien por un «cambio de nivel», bien por un «cambio de tendencia o inclinación». El análisis de la serie temporal puede conducirnos a evaluar el efecto como continuo o permanente, pasajero o momentáneo, instantáneo o retardado.

2. *Diseños de grupo de control no equivalente.* Donde el investigador trabaja con grupos ya formados de una manera natural resultando muy difícil asegurar la equivalencia de los grupos de control y experimental. Se incluyen medidas pre-tratamiento y post-tratamiento, la que nos indica la influencia de la variable independiente; el hecho de que los grupos no sean equivalentes introduce una posible fuente de error en la interpretación. Aunque estos diseños pierden validez interna respecto a los experimentos de campo, su aplicabilidad con grupos naturales les convierte en los diseños más utilizados en la investigación educativa y social.

El método psicométrico

Cuando el objetivo del estudio no es el establecimiento de relaciones de causalidad entre las variables analizadas sino el análisis de la covariación o concomitancia entre las mismas, el investigador puede recurrir a los métodos psicométricos o correlacionales. Su aplicación es procedente cuando no se tiene información sobre los antecedentes del hecho que se estudia y no procede la asignación aleatoria de los sujetos a grupos diferenciados.

El estudio de las relaciones entre variables va a llevarse a cabo en un único momento temporal. El interés del investigador reside más en la búsqueda de regularidades o irregularidades en el conjunto de datos analizados que en el contraste de hipótesis específicas. Su principal contribución consiste en determinar la importancia relativa de las variables para la descripción y explicación de un hecho o fenómeno.

La finalidad, en la mayor parte de los casos, es descriptiva. Se estudia un conjunto de variables (una selección rigurosa de las mismas contribuye a clarificar la interpretación de los resultados) en una muestra con el fin de obtener información sobre el entramado de variables e individuos analizados. Los pasos a seguir en la aplicación de estos métodos serían:

— especificación de la población de interés;
— fijación de la forma de obtención de los datos;
— delimitación de las variables a estudio;
— selección o construcción de los instrumentos para la medida de las variables;
— forma o recogida de los datos (a través de encuestas y cuestionarios);
— establecimiento de las categorías de clasificación de los datos;
— tipo de análisis a realizar, en función del objetivo del estudio.

Este método puede concretarse en diseños muy variados, todos ellos encuadrados en una de estas dos categorías: diseños descriptivos o diseños correlacionales.

1) Diseños descriptivos. Su objetivo fundamental es posibilitar la comparación de comportamientos de grupos en diferentes variables, que no están bajo control del investigador sino que han ocurrido de forma natural y sin intervención experimental. Su finalidad, por tanto, es básicamente descriptiva. Sus limitaciones más importantes son: la falta de control sobre la variable criterio, la imposibilidad de establecimiento de relaciones causales y la distinta familiaridad que los grupos puedan tener respecto a los instrumentos de medida utilizados.

Como modalidad básica de este tipo de diseños destacamos los estudios de encuesta (tipo survey), con un esquema tipo de investigación que incluye las siguientes fases:

— Definición del problema,
— Establecimiento de hipótesis, considerandos,...
— Selección de variables
— Selección/elaboración de instrumentos
— Medida de variables
— Análisis de datos, conclusiones y toma de decisiones

2) Diseños correlacionales. Nos permiten comprender la complejidad del fenómeno o hecho social estudiado, determinando las variables relacionadas con el mismo o las dimensiones básicas en las que se puede descomponer. El esquema tipo de trabajo es similar al presentado para los diseños descriptivos. Pueden distinguirse a su vez tres modalidades:

a) En la modalidad de «diseños de covariación» su finalidad es descriptiva resultan especialmente adecuados para los casos siguientes: realizar análisis exploratorios en campos poco estudiados o con estructuras complejas; establecer tipologías y/o dimensiones básicas en un conjunto de variables; conocer relaciones entre variables, que posteriormente pueden estudiarse con metodologías experimentales.
b) La modalidad de «diseños correlacionales predictivos» se aplican cuando se tiene un mayor conocimiento de la relación entre las variables, de forma que pueden categorizarse como variables criterio o como variables predictoras (este *status* puede cambiar de una investigación a otra). Su objetivo fundamental es predecir, pronosticar, el comportamiento de un grupo de sujetos en una variable criterio determinada, a partir del conocimiento de su relación con un conjunto de variables predictoras.
c) Una tercera modalidad de diseños correlacionales son los denominados «diseños correlacionales causales». Se sitúan en un punto intermedio entre la metodología correlacional y la experimental. Sobrepasan el nivel descriptivo y exploratorio aproximándose al explicativo, aunque para conseguir éste sea necesario que el modelo explicitado resulte de la deducción hipotética de una teoría. No permiten estos diseños generar nuevas teorías, sino contrastar empíricamente las deducciones obtenidas en un marco conceptual. Las relaciones causales se estudiarán a partir de las correlaciones empíricas entre las variables.

El argumento que lo posibilita es el hecho de considerar que la correlación no implica causalidad, mientras que la causalidad implica correlación Batista y Coenders, 2000).

Todos los diseños correlacionales presentados habría que encuadrarlos en la categoría de «estáticos», en la medida en que se analizan las relaciones entre variables o los cambios ocurridos en momentos determinados de tiempo. La inclusión del «tiempo» como variable explicativa nos conduce a la consideración de «diseños dinámicos o evolutivos», de relativa complejidad y con un número importante de modalidades de aplicación: diseños evolutivos descriptivos, estrategias secuenciales, transversales y longitudinales, estrategias secuencial transversa, secuencial de cohorte, secuencial de tiempo..., todas ellas resultantes de la consideración aislada o conjunta de los componentes edad, momento de la medida, cohorte o generación a la que se pertenece.

B) La observación (científica y naturalista)

Esta estrategia metodológica básica de investigación en la Evaluación de Programas debe girar en torno a la observación realizada científicamente; es decir, utilizando técnicas que permitan lograr el máximo control y objetividad de la realidad o hecho educativo, siendo recomendada la utilización de diferentes modalidades metodológicas de observación: sistemática, participativa, externa,...

La observación desempeña un importante papel en la investigación al proporcionar a la ciencia los hechos, uno de sus elementos fundamentales. Mediante la observación, rigurosa y cuidada, el investigador va descubriendo pautas que le capacitan para ir dando forma a sus teorías. La observación científica supone la selección deliberada de algún aspecto significativo de los fenómenos, en cierta situación y en un momento determinado, un detenido examen que puede requerir el uso de procedimientos e instrumentos de precisión y la presentación de resultados.

La observación constituye pues uno de los aspectos más importantes del método científico (de hecho todas las ciencias comienzan por la observación), dependiendo, en gran parte, la consecución de los objetivos científicos del rigor y la precisión con que dicha observación se haya realizado. La observación requiere, para ser científica y en virtud de las condiciones que informan el quehacer científico, ser controlada, o sea, ser objetiva y comprobable; debe emplear una serie de procedimientos y técnicas que eliminen al máximo las posibles fuentes de distorsión y error y establecer adecuadamente las condiciones bajo las cuales se ha observado un determinado fenómeno.

La observación tiene como finalidad obtener información sobre algo, por lo que hay que empezar por delimitar el fenómeno a analizar, situándolo en el marco de una teoría, de la que parte y a la que pretende conformar. La observación debe ser realizada en condiciones específicamente establecidas, las cuales actúan a modo de reglas que establecen la naturaleza de la situación en que se realizarán las observaciones y los procedimientos o instrumentos que deberán ser utilizados.

En el contexto de la Evaluación de Programas, la observación científica se concreta como método que podemos denominar «observación naturalista». Este método se caracteriza porque el análisis de la realidad se basa en la observación de los comportamientos naturales del grupo social de referencia: clase, barrio, comunidad, distrito... Su objetivo fundamental es el estudio intensivo de los fenómenos complejos que conforman la vida de un grupo de personas, con finalidad comprensiva que debe conducirnos a propuestas para subsanar las carencias educativas y culturales percibidas.

Existen numerosas técnicas de observación, que se generan en función del ambiente en que se realiza (ambiente natural y ambiente artificial), en función de la utilización o no de registros observacionales (observación estructurada o no estructurada), y en función de la actitud que adopte el investigador respecto al proceso social observado (investigación participativa o no participativa).

El diseño determina la posibilidad de detectar y comprender los procesos y el impacto de un programa educativo, dada la multiplicidad de factores que influencian su operación y resultados (De la Orden, 1990). Podemos considerar las siguientes pautas como esquema general de trabajo en la aplicación del método observacional (Martínez Arias, 1981; Garaigordobil, 1999):

— definir las metas del estudio
— establecer las muestras que serán observadas
— decidir si la observación será participante o no
— modo de entrar en contacto con los sujetos
— muestrear los ambientes y situaciones en que se desarrollará
— muestrear las unidades temporales
— establecer el procedimiento de recogida de datos
— fijar las categorías de clasificación de las conductas a observar
— considerar las posibles técnicas de organización y análisis de datos

Las ventajas fundamentales residen en la posibilidad de estudiar conductas no verbales y estudiar comportamientos en toda su complejidad. Sus principales inconvenientes para una aplicación optimizada serían la carencia de información sobre variables antecedentes que puedan ayudar a describir los hechos observados; los riesgos de carga subjetiva que pueden asociarse a las conclusiones obtenidas por el observador; los riesgos de reactividad de la presencia del observador (pueden evitarse con una presencia larga o encubierta); la dificultad de establecer categorías claras que orienten los registros; la utilización de muestras reducidas; el entrenamiento de los distintos observadores; y las dificultades del investigador para introducirse en el ambiente que se desea observar (Rodríguez, Gil y García Jiménez, 1996; Goetz y LeCompte, 19888; LeCompte, 1995).

La observación como técnica básica en la investigación cualitativa puede tener aplicaciones en diferentes formas y contextos, lo que determina las diferentes estrategias de investigación cualitativa, con notas peculiares y diversidad de concreciones:

estudio de casos, de grupos, estudios etnográficos (Rodríguez, Gil y García Jiménez, 1999).

C) El entorno interpretativo-crítico

La elaboración de un diseño desde la perspectiva crítica implica aceptar las siguientes premisas básicas (Colás y Rebollo, 1993, p. 112):

— El diseño está estrechamente vinculado, conceptual y prácticamente, con la acción-intervención
— Los receptores del programa se convierten en agentes de la evaluación
— Los procesos evaluativos los realizan y sirven a los implicados
— La secuencia evaluativa es cíclica: comienza con el diagnóstico y termina en la valoración global del contexto social y personal. No se pueden aislar aspectos o dimensiones del programa para su evaluación.
— El análisis e interpretación de los datos de hace de forma colectiva a la luz de un contexto social y político más amplio. En el análisis se utiliza información sobre cómo, quién, cuando y dónde se obtienen los datos

La investigación en la acción deben ser entendida como la síntesis de dos dimensiones indisociables de un mismo proyecto; sólo así se puede comprender la práctica; la finalidad de la investigación es a la vez explicativa y normativa: se investiga para perfeccionar nuestro conocimiento, fundamentar racionalmente la práctica y transformar la realidad. La actividad desarrollada es a la vez un acto de investigación y un acto sustantivo.

La investigación-acción podría definirse como el estudio de una situación social (escolar) con el fin de mejorar la calidad de acción dentro de la misma. En la investigación-acción las teorías no se hacen válidas de formas aislada, para ser aplicadas después a la práctica. Se hacen válidas por medio de la práctica (Rodríguez Rojo, 1991).

La investigación-acción propuesta por Kurt Lewin (1951) posee las siguientes características:

— Es una actividad desarrollada por grupos o comunidades con el propósito de cambiar las circunstancias de acuerdo con una idea compartida.
— Es una práctica social reflexiva, donde no cabe hacer ninguna distinción entre la práctica que se investiga y el proceso de investigación de esa práctica. Las prácticas (sociales, escolares,...) son consideradas como «actos de investigación», como «teorías en acción» o «explicaciones hipotéticas que han de ser examinadas y valoradas en términos de su potencialidad para realizar cambios valiosos. Si una práctica social como la enseñanza es considerada como una actividad reflexiva ya no existe la división del trabajo entre los que la ejercen y los que la investigan (Elliot y otros, 1986, p. 12; Elliot, 1990; Elliot, 1993).

Lewin establecía un proceso disciplinado de investigación en la acción, con un claro paralelismo con el método científico, y articulado en una espiral de actividades realizadas en el siguiente orden:

— clarificar y diagnosticar una situación problemática para la práctica
— formular estrategias de acción para resolver el problema
— poner en práctica y evaluar las estrategias de acción
— nueva aclaración y diagnosis de la situación

Los puntos esenciales que definen y caracterizan la investigación-acción y el papel a desempeñar por esta estrategia metodológica de investigación en los procesos de innovación son analizados con detalle por Pérez Serrano (1990, pp. 53-54).

La investigación participativa representa una propuesta metodológica para el cambio social. Es una perspectiva científica o ideológica para estimular, apoyar y facilitar los procesos de transformación. Sus principios epistemológicos significan una manera de acercamiento a la realidad social: se trata de conocer transformando (Gabarrón y Hernández, 1994, p. 23).

BIBLIOGRAFÍA

Alvira, F. (1986): «Diseños de investigación social: Criterios operativos» en F. Alvira, M. García Ferrando y J. Ibáñez *Análisis de la realidad social.* Madrid, Alianza Editorial.

Alvira, F. (1991): *Metodología de la evaluación de programas.* Cuadernos Metodológicos nº 2, Madrid, CIS.

Amezcua, C. y Jiménez, A. (1996): *Evaluación de Programas Sociales.* Madrid, Díaz de Santos.

Anguera, M.T. (1990): «Programas de intervención. ¿Hasta que punto es factible la evaluación?». *Revista de Investigación Educativa,* nº 16, pp. 77-93.

Bartolomé, M. (1990): «Evaluación y optimización de los diseños de intervención». *Revista de Investigación Educativa,* nº 16, 39-60

Batista, J.M. y Coenders, G. (2000): *Modelos de ecuaciones estructurales.* Madrid, La Muralla-Hespérides.

Carballo, R. (1990): *Evaluación de un Programas de Formación Ocupacional.* Tesis Doctoral. Madrid, Universidad Complutense.

Colás, P. y Rebollo, Mª A. (1993): *Evaluación de programas. Una guía práctica.* Sevilla, Kronos.

Cronbach, L.J. (1982): *Designing Evaluations of Educational and Social Programs.* San Francisco, Jossey-Bass.

De la Orden, A. (1990): «Evaluación de los efectos de los programas de intervención». *Revista de Investigación Educativa,* nº 16, 61-76.

Elliot, J. y otros (1986): *Investigación en el aula.* Generalitat Valenciana, Valencia.

Elliot, J. (1990): *La investigación-acción en educación.* Madrid, Morata.

Elliot, J. (1993): *El cambio educativo desde la investigación-acción*. Madrid, Morata.
Finnegan, J.R. y otros (1988): «Measuring and tracking education program implementation: The Minnesota Heart Health Program». *Health Educational Quarterly*, 16 (1), 77-90.
Fitz-Gibbon, C.T. y Morris, L.L. (1978): *How to design a Program Evaluation*.
Gabarrón, L.R. y Hernández Landa, L. (1994): *Investigación participativa*. Cuadernos Metodológicos. Madrid, CIS.
Garaigordobil, M. (1999): «Técnicas observacionales en la construcción de instrumentos de evaluación psicológica y en la evaluación continua de programas». En M.T. Anguera *Observación en el aula: Aplicaciones*. Barcelona, Ed. Universitat de Barcelona, Monografía de psicología.
Goetz, J.D. y Lecompte, M.D. (1988): *Etnografía y diseño cualitativo en investigación educativa*. Madrid, Morata.
Guba, E.G. (1985): «Criterios de credibilidad en la investigación naturalista» en J. Jimeno y A. Pérez (eds.) *La enseñanza: su teoría y su práctica*. Madrid, Akal.
Kerlinger, F.N. (1977): «The influence of Reserach on Education Practice». *Educational Researcher*, Vol. 6, nº 8, pp. 5-12.
Lecompte, M.D. (1995): «Un matrimonio conveniente: diseño de investigación cualitativa y estándares para la evaluación de programas». *Revista Electrónica de Investigación y Evaluación Educativa*, I, 1.
Lewin, K. (1951): *Field theory in Social Science*. New York, Harper.
Martínez Arias, R. (1981): «Diseños de investigación en psicología ambiental». En A. Jiménez Burillo *Psicología y medio ambiente*. Madrid, Ceotma, MOPU.
Martínez Mediano, C. (1997): «La teoría de la evaluación de programas». En AIDIPE *Actas del VIII Congreso de Modelos de Investigación Educativa*.
Mayer, R.R. (1985): *Policy and program planning*. Englewood Cliffs, N.J., Prentice Hall.
Pérez Carbonell, A. (1998): *Metodología de la Evaluación de Programas: Evaluación de Programas de Postgrado en la Universidad de Valencia*. Tesis Doctoral. Valencia, Universidad de Valencia.
Pérez Juste, R. (1995): «Evaluación de programas educativos». En L.M. Villar y A. Medina *La evaluación de programas educativos, centros y profesores*. Madrid, Universitas.
Pérez Serrano, G. (1990): *Investigación-acción. Aplicaciones al campo social y educativo*. Madrid, Dykinson.
Rodríguez, G., Gil, J. y García Jiménez, E. (1999): *Metodología de investigación cualitativa*. Málaga, Aljibe.
Rodríguez Rojo, M. (1991): «Cómo aprender qué es una investigación acción». *Investigación en el Aula*, 13, 59-66.
Sanders, J.R. (1998): *Estándares para la evaluación de programas (Comité Conjunto de Estándares para la Evaluación Educativa)*. Bilbao, Mensajero.
Scriven, M.S. (1973): «Goal-free evaluation» en E.R. House *School Evalutions: The politics and process*. Berkeley, California, McCutchan.
Tejedor, F.J. (1990): «Perspectiva metodológica del diagnóstico y evaluación de necesidades en el ámbito educativo». *Revista de Investigación Educativa*, nº 16, 15-38.

Tejedor, F.J. (1994): «La experimentación como método de investigación educativa», en V. García Hoz *Problemas y métodos de investigación en educación personalizada,* pp. 256-285. Madrid, Rialp.

Tejedor, F.J., García-Valcárcel, A. y Rodríguez, Mª J. (1994): «Perspectivas metodológicas de la evaluación de programas en el ámbito educativo». *Revista de Ivestigación Educativa,* nº 24, pp. 93-128.

Weis, C.H. (1982): *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción social.* México, Trillas.

Witkin, B.N. (1984): *Assessing needs in Educational and Social Programs.* London, Jossey-Bass Publ.

336

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 341-356

LA VALIDEZ EN LA EVALUACIÓN DE PROGRAMAS

J. M. Jornet, J. M. Suárez y A. Pérez Carbonell[1]

INTRODUCCIÓN

El tema de la Validez, en cualquiera de los ámbitos que afectan a las CC. De la Educación, es un tema recurrente y, por lo tanto, tratado con exhaustividad y acierto por la gran mayoría de autores que, en algún momento, lo han atendido. Por ello, más que acercarnos a él desde una perspectiva estrictamente de fundamentación metodológica, en la que poco se puede aportar, estimamos que quizá es más adecuado revisar este tema desde las implicaciones que se derivan de las experiencias evaluativas en las que hemos estado inmersos, refiriendo eso sí los componentes teórico-metodológicos con los que nos sentimos más identificados. Por ello, este trabajo lo estructuramos en dos grandes apartados: a) Del concepto de Validez, y b) Estrategias de consecución de la Validez.

1. DEL CONCEPTO DE VALIDEZ

El Concepto de Validez es dependiente del proceso a que se aplique, así como de las características de la posición epìstemológica desde la que se aborde. Por ello, estimamos necesario antes de referirnos al Concepto de Validez, definir nuestra posición respecto a la Evaluación de Programas, como proceso, y en cuanto a su orientación epistemológica y metodológica.

La *Evaluación de Programas*, como Investigación Aplicada, viene especificada fundamentalmente por su objeto, su finalidad o finalidades así como por el contexto. Como proceso de investigación, éste debe ser riguroso, estar bien controlado y realizarse de forma sistemática. Entendemos, pues, que la Evaluación de Programas (Alvira, 1996; Pérez Carbonell, 1998 y 1999; Rogers y Owen, 1999; Pérez Juste, 2000):

1 Dpto. MIDE de la Universitat de València. Fac. Filosofía y CC. De la Educación. Avda. Blasco Ibáñez, 30. Tél. 963864430. e-mail: jornet@uv.es, rodrigus@uv.es, pereza@uv.es

- Requiere del mismo grado de control, rigor y sistematización que se le exige a cualquier otro tipo de investigación que pretende alcanzar un conocimiento, máxime teniendo en cuenta que pretende identificar elementos para intervenir sobre lo evaluado (es decir, de una u otra forma, pretende servir de base para un proceso en que se implican decisiones).
- Constituye un tipo de investigación con características muy definidas:
  - — *Tiene una utilidad inmediata* que exige la aportación de un alto grado de confianza en los resultados que de ella se ofrecen así como en los procedimientos utilizados para alcanzar dichos resultados.
  - — *Esta impregnada de cuestiones de valor*, como ¿conviene modificar algo?, ¿estamos realizando un trabajo efectivo?, ¿se esta alcanzado lo pretendido?...
  - — *Se encuentra condicionada por factores sociales y políticos* que la involucran en situaciones excesivamente complejas.
- Por ello, precisa de una *Planificación* en la que se anticipe el comportamiento de los componentes / elementos que intervienen en su propio proceso.

No obstante, bajo el término *Evaluación de Programas coexisten diversos tipos de aproximaciones, que resultan en ocasiones muy divergentes en cuanto a los planteamientos epistemológicos y metodológicos* en que se sustentan, o bien, simplemente, en cuanto a sus campos de actuación. En este caso, la reflexión que aportamos se centra respecto a la Evaluación de Programas como campo metodológico, es decir, asumiendola como sinónimo de la Investigación Evaluativa.

Si realizamos un *breve recorrido histórico* sobre las corrientes que han estado presentes en el panorama de la Investigación Educativa, y que por lo tanto influyen en la delimitación del concepto de Validez en la Evaluación de Programas, hay que señalar que ha sido el positivismo y postpositivismo con su propio método científico el que se presenta con una mayor tradición científica dentro de las Ciencias de la Educación. Sin embargo, en esta tradición la aplicación de estas aproximaciones no ha sido uniforme y se puede observar un continuo que va desde el cumplimiento absoluto de todos los aspectos que marcan y caracterizan un planteamiento científico-naturalista a nivel epistemológico, ontológico y metodológico, hasta el cumplimiento más laxo de los mismos (De Miguel, 1988; Suárez, 1989; Dendaluze, 1995, 1997; Aliaga, 2000). Esto quiere decir, que si bien en la metodología Positivista y Postpositivista hay una serie de requisitos que deben estar presentes para llegar a estimar el impacto de un posible agente causal (variable Programa) en algún efecto potencial (variable Resultado al implementar el Programa), como son el control y la aleatorización, lo están pero con graduaciones que van desde un mayor a un menor cumplimiento. Hablamos, trasladándolo a la terminología que utilizamos, del continuo que va desde lo Experimental a lo No Experimental pasando por lo Cuasi-experimental.

En el momento en el que el positivismo (años 60 hasta años 80) recibe mayores críticas es cuando empieza a tomar fuerza, en cuanto a metodología de investigación en Educación, el *Constructivismo*. Como es sabido, este enfoque está integrado por una gran variedad de fuentes y que confluyen en torno al concepto de *Versthen* (comprensión de los significados de las acciones humanas, Jacobs, 1987). Su finalidad es la de

llegar a comprender la intencionalidad en la actuación de un agente y no si ese agente es la causa o el efecto. Desde esta perspectiva, la explicación causal y la predicción del positivismo, es sustituido por la Comprensión —de la intencionalidad de las acciones y de las percepciones de los sujetos que pueden contemplarse desde la Investigación fenomenológica, la etnográfica, la participativa, la Investigación-Acción, la fenomenográfica...—.

Superado el debate paradigmático (Cook y Reichardt, 1986; Cook y Shadish, 1994; Reichardt y Rallis, 1994), en los últimos años los paradigmas *Cualitativos*, y en especial el *Interpretativo* y para el *Cambio*, influyen de forma decisiva en todos los acercamientos que pueden enmarcarse en la Investigación Educativa, y, con mayor énfasis, en sus ámbitos más aplicados (Colás, 1995; Chelimsky y Shadish, 1997). Las posiciones actuales más representativas abogan por la *Complementariedad Metodológica* (Krathwohl, 1987; Mateo, 1990; Tejedor, García-Valcárcel y Rodríguez, 1994; Dendaluze, 1995; Creswell, 1994, 1998; Bericat, 1998; Buendía, 1998). Lo cierto es que en el ámbito de la Evaluación de Programas, es difícil dar respuesta ajustada a los problemas que aborda, si no nos situamos en una posición flexible. Desde ella nos aproximaremos a la revisión del Concepto de Validez.

Desde este punto de vista, en la Evaluación de Programas se pueden identificar diversas *facetas del concepto de Validez* de especial énfasis. Para acercarnos a ellas partimos de una definición genérica de la Validez en la Evaluación de Programas:

> *El grado en que un proceso de Evaluación de Programas atiende de forma adecuada los componentes metodológicos implicados con el fin de:*
>
> - *representar de forma realísta las características del objeto que evalúa,*
> - *facilitar la consecución de los fines establecidos para el Plan de Evaluación,*
> - *respetar los intereses de todos los implicados o afectados por la evaluación,*
> - *representar el contexto en el desarrollo del Plan de Evaluación, en la modulación de la interpretación de resultados y en la adecuación de las consecuencias de la evaluación, y*
> - *coadyuvar a la mejora o innovación de lo evaluado.*

En definitiva, la *Validez es un concepto multidimensional*, que ha sido referido en diversos momentos y situaciones haciendo énfasis en diferentes acepciones. Aplicado —y aplicable— a cualquier estrategia de acercamiento a la realidad (desde la investigación básica, los instrumentos de recogida de información o la investigación aplicada y el diagnóstico), es el criterio sustantivo desde el que valorar la calidad de cualquiera de estos procesos. Desde la posición que vamos desarrollando, la Validez alude entonces al *Criterio de Calidad de todo el proceso, y como tal es relativo* y por lo tanto gradual. No se puede hablar de la *Validez* como una característica que se dé o que esté ausente, sino que, en todo caso, intrínsecamente es cualitativo y globalizador, y está necesariamente vinculado a las características del objeto de evaluación, la finalidad y uso de los resultados, etc. (Swanborn, 1996; Perales 2000). Por ello, vamos a intentar aproximarnos a él, analizando diversas acepciones que le serían aplicables. Como medio de aproxima-

ción, y con el fin de poder establecer un orden en el análisis y exposición del concepto utilizamos una estrategia basada en el Análisis de Facetas de Guttman. Así, distinguimos en el concepto *dos tipos de Facetas*:

- las que responden a los componentes clásicos metodológicos de interpretación de la Validez, como referencias conceptuales desde las que abordar la reformulación del concepto, y a las que denominamos *Facetas subyacentes metodológicas de la Validez*,
- las que se pueden identificar en los componentes específicos de un proceso de Evaluación de Programas, que denominamos *Facetas adaptativas de la Validez*.

De acuerdo con lo expuesto anteriormente, pasamos a comentar las siguientes *Facetas de la Validez*:

a) *Facetas subyacentes metodológicas de la Validez*, nos referimos a aquellas que se pueden entender como compomentes implicados en cualquier proceso de investigación, y que se especificarán en cuanto a su importancia o relevancia, atendiendo a las características concretas de dicho proceso (definidas por el objeto a investigar, la finalidad, etc...). Entre ellas, se pueden identificar los siguientes tópicos (Rutman, 1984):

- *Facetas genéricas*: son aquéllas que se pueden identificar con mayor o menor énfasis, dependiendo del tipo de situación (Programa, Plan de Evaluación, etc...).
  1. *Validez Interna:* se trata de demostrar que la asociación constatada entre el Programa implementado y el efecto medido se debe a una relación causal. Desde esta posición, la entendemos como *conjunto de criterios de calidad que se orientan por los principios de consistencia y credibilidad de los diferentes elementos que componen el proceso de evaluación.*
  2. *Validez Externa:* se trata de poder generalizar a constructos de mayor nivel teórico o amplitud partiendo de tratamientos y medidas específicas. Entendida como conjunto de criterios de calidad. Integra el *conjunto de criterios de calidad que se orientan por el principio de generalidad*: conseguir una adecuada pluralidad en los diferentes elementos que componen el proceso de evaluación *consistente* con la definición asumida en cada caso.
- *Facetas específicas*: son aquéllas que se pueden identificar con mayor o menor énfasis, dependiendo del tipo de situación (Programa, Plan de Evaluación, etc...).
  3. *Validez de Constructo:* se trata de realizar inferencias acerca de las construcciones teóricas que han estado orientando el planteamiento de la evaluación (por ejemplo, teorías organizacionales, en el caso de evaluaciones institucionales; o teorías instruccionales, en casos de programas de instrucción...). Respetando la ausencia de teorías formalmente construídas y adecuándose al nivel de desarrollo teórico existente en cada caso. En este sentido, hay que señalar que en el ámbito educativo, los procesos de evaluación de programas

se enfrentan a situaciones muy diversas en las que se pueden integra, a su vez, variables e indicadores de muy diversa construcción teórica. En este caso, se trata de adecuar el proceso metodológico atendidendo al grado de desarrollo existente para cada caso, tratando cada situación en virtud del desarrollo teórico existente. Efectivamente, son muchas las situaciones en las que no existe no ya una teoría consolidada que pueda orientar en el proceso, sino la debilidad de la mayoría de las existentes, o la ausencia de teorías para gran cantidad de situaciones. No obstante, la Evaluación de Programas, debe ofrecer respuestas en cualquier situación, si bien no debe forzar el desarrollo de sus procesos en virtud de requerimientos metodológicos que se han podido desarrollar en el marco de usos guiados por teorías. Así, el planteamiento metodológico deberá adaptarse a las ausencias o debilidades teóricas, radicando en este respeto a la realidad de la representación, precisamente, la validez de constructo del proceso evaluativo. Finalmente, hay que destacar el potencial de la Evaluación para aportar conocimiento —y, por tanto, para facilitar información para la construcción teórica— (De Miguel, 1999).

4. *Validez de Conclusión:* se trata de realizar inferencias válidas acerca de si un Programa y su posible efecto están relacionados, aún sin saber si esa relación es o no causal. Las Conclusiones, en este caso, deben identificarse en relación con los fines declarados en la Evaluación. Así, deben cumplir necesariamente los principios de consistencia y generalidad, atendiendo a la *relevancia y utilidad* de la información derivada de la evaluación.

No obstante, como señalamos anteriormente, estas facetas constituyen referencias globales en cuanto al planteamiento metodológico para cualquier tipo de Evaluación de Programas. La importancia diferencial de cada una de ellas, dependerá, como es obvio, del tipo de Programa a Evaluar, así como de los demás factores implicados, fundamentalmente la finalidad de la evaluación y su contexto, y condiciones de desarrollo. En este sentido, únicamente señalar algún ejemplo:

- Durante el Diseño de un Programa de Formación se realiza una evaluación para comprobar su eficacia, eficiencia y funcionalidad[2]; el objetivo es identificar los elementos a mejorar y, en definitiva, optimizar el Programa. En este supuesto, pueden concurrir todas las facetas descritas.
- Por otra parte, podemos acercarnos a otro supuesto: la evaluación del grado en que un Programa de Formación Ocupacional, orientado a la inserción laboral de las personas que asisten a él, logra sus objetivos. En este caso, la finalidad del Programa de Formación (inserción laboral), así como las condiciones de control en que se desarrolla, sitúan el énfasis en otros aspectos de la Validez, como la calidad en la representación del logro de los objetivos que se perseguían en su puesta en marcha, o la utilidad de la información que de él se deriva para aquéllos que deben tomar decisiones acerca del Programa. El impacto del Programa,

2 En el sentido definido por De la Orden (1990).

por tanto, sitúa el interés de la Evaluación en la esfera de sus consecuencias, más que en la de sus procesos. En esta línea, la evaluación de este tipo de Programas puede orientarse primordialmente hacia el impacto del programa, por lo que la Validez de la Evaluación, también se dirigirá hacia este aspecto (Mohr, 95; Apodaka, 99). Lo más importante en este caso será que la Evaluación pueda ofrecer información adecuada acerca de la inserción profesional de los asistentes al Programa. La amplitud del concepto que reflejen los indicadores de inserción, así como los elementos explicativos de la misma, constituirán elementos clave para la validación del plan de evaluación. Así, por ejemplo, no se podrá considerar igualmente válido un Programa que se limite a controlar cuántas personas han encontrado trabajo después de realizar el Programa de Formación (en diversos momentos: después de un mes, tres meses o un año...), que otro que además considere otros elementos, como la adecuación del trabajo conseguido en función de la formación recibida, la covariación posible de la inserción con otras variables (personales —edad, sexo, curriculum...—, sectoriales y/o sociales —tejido empresarial de la localidad en que se encuentra trabajo, elementos cíclicos de la economía, ...—, etc...), o el análisis de los beneficios personales obtenidos a partir del Programa (con independencia de la mera inserción).

Estas matizaciones nos llevan a considerar otras facetas de la Validez de la Evaluación de Programas, que cobran especial énfasis en virtud de las características de la Evaluación de Programas como ámbito aplicado de investigación. A ellas nos referimos en el apartado siguiente.

b) *Facetas adaptativas de la Validez*, porque se implican en tanto en cuanto el proceso de Evaluación de Programas debe atender a todos los componentes necesarios para conseguir sus fines (mejora, innovación, cambio...). Se pueden identificar en los siguientes grandes núcleos procesuales:

- *En relación con la Representación del Objeto.* Se trata del grado en que, en la Evaluación, se seleccionan indicadores y criterios que se ajusten a aquello que se desea evaluar. La identificación de instrumentos, fuentes de información, tipo de información que se va a facilitar, utilización de la misma, etc..., forman parte de la misma.

  En este sentido, hay que señalar que en la Evaluación de Programas, en múltiples ocasiones, se producen disfunciones entre los planteamientos que animan los procesos de evaluación y su traducción u operativización en los indicadores que los sustentan. Así, revisemos únicamente algún ejemplo: los procesos de mejora de la calidad aluden a este concepto como la referencia que orienta la evaluación, así como la finalidad de la misma (mejora de la calidad). No obstante, las definiciones acerca de la Calidad de cualquier programa educativo suelen ser muy genéricas. Recuerdan en alguna medida el rol que juegan las definiciones de Rasgos Psicológicos como orientadoras de los tests psico-

métricos[3]. De esta forma bajo una misma definición de Calidad de un Programa se pueden identificar evaluaciones muy diversas, orientadas hacia el Programa, o hacia su efecto, o únicamente sobre algunos elementos procesuales[4] (Dendaluze, 1991).

En la Evaluación de Programas, la *representación del objeto, necesariamemte, nace del compromiso entre todos los elementos de definición de la evaluación*. No sólo se trata de definir qué se evalúa, sino que hay que entender que ello es necesariamente interdependiente con todos los restantes componentes de la evaluación: para qué, cómo, a quién y de qué modo se pide información, qué información se va a aportar, qué usos se harán con ella, etc... Este compromiso global de todos los componentes en la definición del objeto a evaluar persigue la idea de *Coherencia* como base para aspirar a la Validez en cualquier proceso de evaluación. Por ello, es un aspecto central en la Planificación y diseño de la Evaluación.

Implica por lo tanto, todas las facetas que se relacionan con el planteamiento metodológico e instrumental: establecimiento del diseño, determinación de los procesos de la evaluación, selección de indicadores e instrumentos, tipos de análisis de la información, usos de la misma, etc...

- *Adaptación al proceso evaluado*. Se trata de llevar a cabo los diferentes elementos del proceso de la Evaluación, respetanto el proceso natural de desarrollo del Programa Educativo. En esta línea, hay que señalar los elementos que coadyuvan a la consecución de la Validez, se pueden sintetizar en el *Respeto* a los procesos que se evalúan, no interfiriendo en los mismos. La introducción de expectativas por el hecho de la realización de la evaluación, los condicionantes que pueda suponer su realización para el proceso o para los diferentes actores implicados en el mismo, entre otros aspectos, pueden introducir elementos que interfieran en el normal desenvolvimiento de lo que se está evaluando.
- *Integración Contextual*. Otro elemento sustancial de la Validez es la consideración de las diferentes variables que pueden afectar al Programa evaluado y, por ello, deben ser atendidas como elementos de contextualización. Estos elementos servirán de base para la extensión de la Comprensión, y en su caso, Explicación de los resultados. Por ello, son fundamentales como componentes de modulación de la interpretación y comprensión de la situación evaluada.
- *Participación e implicación de audiencias*. Desde la posición epistemológica y metodológica en que nos situamos, es necesario considerar que la consecución de los fines de la evaluación sólo son posibles si se atiende adecuadamente la Partici-

---

3 Recuérdese, por ejemplo, la siguiente definición: el test ... está destinado a valorar la capacidad para conceptualizar y aplicar el razonamiento sistemático a nuevos problemas y aprecisar las funciones centrales de la inteligencia: abstracción y comprensión de relaciones. Según los autores del mismo es el objetivo del Test D-48. La excesiva distancia entre la definición del objeto a evaluar (Rasgo: Inteligencia), y la operativización de su medida es un claro indicador de falta de Validez (en este caso, hablaríamos de carencias en cuanto a su Validez de Constructo y de Contenido, como instrumento de medida).

4 En este mismo número monográfico hacemos, junto a la profesora Mª J. Perales, una revisión de los Planes de Evaluación de la Formación Ocupacional y Contínua. En él se incluye una referencia concreta a la consideración del concepto de Calidad en estas Evaluaciones. Puede ilustrar los extremos que en este punto comentamos.

pación de los diferentes actores implicados en el Programa (De Miguel, 1995). La Validez de la Evaluación, desde esta perpectiva, dependerá del grado en que se haya ajustado la participación de todos ellos, de acuerdo con su nivel de implicación y expectativas. Estos aspectos se conjugan en diferentes momentos del proceso, desde la participación en la definición del objeto a evaluar y sus finalidades, pasando por el tipo de información que se requiere y cómo se recoge, hasta la información que se hace disponible y cómo se difunda. Lo cierto es que impregna todo el proceso de evaluación y, por ello, debe ser adecuadamente atendida en toda su Planificación y Desarrollo.

- *Compromiso con el Cambio: Credibilidad y Utilidad.* Finalmente, esta faceta es probablemente la más determinante en cuanto a las consecuencias de la Evaluación. La Evaluación de Programas, desde la posición de que partimos, sólo es posible considerarla válida si aporta los elementos de mejora e innovación que se pretendía identificar y que justificaban la puesta en marcha del proceso. Estas facetas, ya descritas por otros autores (Tejedor, 2000; Mateo, 1990), son a su vez consecuencia de haber atendido adecuadamente todas las anteriormente descritas, y base para la valoración global del proceso. En definitiva, procesos evaluativos metodológicamente bien construídos y desarrollados, no son asumidos como válidos, por la falta de credibilidad que dimana de alguno de los componentes anteriores que no se han atendido bien, o por la falta de utilidad de la Evaluación.

  En cualquier caso, esta faceta no depende de los momentos finales del proceso evaluativo, sino que empieza a construirse en el momento en que se comienza con la evaluación, adaptando de forma coherente todo el proceso y atendiendo todas las facetas señaladas.

2. ESTRATEGIAS DE CONSECUCIÓN DE LA VALIDEZ EN EVALUACIÓN DE PROGRAMAS

En los procesos de investigación y evaluación enmarcables dentro de posiciones postpositivistas, la Validez —tanto Interna como Externa— se vincula a diversos factores, favorables a su consecución, o bien que atentan contra ella. Desde la posición que revisamos, aunque constituye una reinterpretación conceptual, obviamente también los hay. Sin embargo, dadas las características de esta presentación, no vamos a realizar un análisis exhaustivo de los mismos, sino que vamos a acercarnos a su delimitación tomando como base las diversas etapas de desarrollo en que podría enmarcarse un proceso de Evaluación de Programas[5]. Para ello, integramos las acciones que se implican

5 Identificar *Etapas* alude a la referencia de una secuencia lógica que nos permita centrar el análisis. Entendemos que en estos Procesos debe adoptarse una actitud dinámica que los convierta en un esquema de trabajo flexible. Ello no quiere decir que queden fuera de este planteamiento situaciones como los diseños emergentes. Es obvio que, en ellos, no sería posible identificar estas grandes categorías de funcionamiento, tal cual lo hacemos aquí. Sin embargo, la elección del esquema de reflexión se basa en la pretensión de abordar el tema desde una perpectiva general, aunque produzca problemas en cuanto a su aplicación a tipos de aproximaciones más específicas. El fundamento del enfoque, no obstante, es el mismo.

en los procesos de evaluación en dos grandes conjuntos: a) la Planificación de la Evaluación, y b) la Gestión de la Información. Pasamos a comentarlas a continuación.

- *Planificación de la Evaluación: Construcción del plan de trabajo, adaptación y desarrollo del mismo.*

  Por *Planificación* entendemos el conjunto de actividades que van anticipando y haciendo posible todos los elementos a tener en cuenta en el proceso de evaluación (Jornet, Suárez y Belloch, 1998). Incluirá desde los elementos metodológicos clásicos (decisiones en cuanto al diseño de la evaluación, elementos de operativización, etc...), junto a todos los componentes psicosociales implicados en el desarrollo de una evaluación (identificación de audiencias, promoción de la aceptación de la evaluación, establecimiento de consensos en cuanto a todos los componentes del proceso, etc...).
  En cualquier caso, debe tenerse en cuenta que hablar de Planificación no implica necesariamente referirse a procesos en que es posible anticipar todos los elementos a tener en cuenta. Por ello, si bien se puede entender como una acción que comienza con anterioridad a la puesta en marcha de cualquier evaluación, debe abordarse como un componente de flexibilización, que debe presidir todo el desarrollo del proceso, facilitando las modificaciones y adaptaciones del mismo según sea necesario.
  Por ello, los elementos que pueden coadyuvar en esta actividad a la consecución de la Validez son muy diversos. Entre ellos los más relevantes pueden ser:

- *Participantes, roles y responsabilidades.*

  Probablemente uno de los elementos clave de cualquier proceso de evaluación es la integración de los diversos implicados de acuerdo con su grado de responsabilidad y rol. Obviamente, desde la definición de lo que va a ser objeto de evaluación, pasando por la finalidad de la misma y uso de la información, etc...la Evaluación de Programas debe estar enmarcada en un proceso claro de participación, facilitando el desarrollo de evaluaciones en los que todos los elementos estén consensuados. Es claro que en este tipo de procesos nos encontramos con situaciones muy diversas, desde evaluaciones orientadas a la rendición de cuentas en las que se sustenta el proceso sobre evaluadores externos hasta otras netamente orientadas a la mejora e innovación basadas en la autoevaluación. En cualquier caso, en todas ellas, habrá que analizar los grados de participación de todos los colectivos implicados y favorecer su participación, promoviendo estrategias de evaluación basadas en la *aceptación y credibilidad*.
  En este marco, son importantes las estrategias basadas en la creación de comités que integren a representantes de los diversos colectivos implicados, determinar el rol de los evaluadores en los mismos. Estas estrategias pueden constituir la base para la consecución del consenso en los diferentes momentos por los que transita un proceso de evaluación.

- *Definición del Objeto a Evaluar y Finalidad de la Evaluación*

El objetivo inicial es delimitar el objeto y finalidades de la Evaluación en términos de *validez/credibilidad*, tanto *interna* —la consistencia del planteamiento del problema respecto a la situación específica que se pretende abordar y a las otras respuestas que se han dado en situaciones similares—, como *externa* —credibilidad para los agentes y audiencias implicadas en los mismos—.
En este sentido, las estrategias básicas de acercamiento a la *Validez*, requieren, junto a un planteamiento metodológico riguroso y ajustado, de la *participación de todos los implicados*. Se trata de valorar y adecuar la orientación de la evaluación, identificar los obstáculos que puedan producirse para llevar a cabo el proceso (desde su realización a sus consecuencias), promover la aceptación de la evaluación y anticipar los limites de utilización de la Evaluación. Para ello, la *negociación* constituye la base más adecuada para desarrollar este proceso.
Los componentes básicos en este caso son:

— Determinar qué es lo que se va a evaluar a nivel operativo, y
— Definir los usos finales de la evaluación.

En este caso, un elemento que se va introduciendo paulatinamente es el de *Contrato de Evaluación*. Una fuente frecuente de invalidez de los procesos de Evaluación de Programas reside en los cambios que se van introduciendo a lo largo del proceso de evaluación, que pueden llevar a conllevar utilizaciones no previstas de la misma. Asimismo, en muchos casos hemos observado como la institucionalización de un proceso en lugar de constituir un elemento de mejora del mismo, conlleva perjuicios en los usos de los sistemas de evaluación; la presencia de cambios sucesivos en los gestores de la institución usuaria de la evaluación, facilita este tipo de cambios[6].

- *Construir el plan de trabajo*

Se trata de traducir en términos de actuación el problema que se ha planteado. Esta traducción que guía todo el plan de trabajo debe reunir condiciones de *fidelidad y generalidad*. Así, por una parte, es preciso mantener una conexión permanente con las ideas clave o características fundamentales de la definición del

6 Obviamente, además de un problema de validez para las Evaluaciones de Programas, constituye un problema ético para los evaluadores. Lo cierto es que en este contexto, es claro que si un sistema se desarrolla para una institución, atendiendo a un uso determinado, si bien el sistema es de la Institución, los implicados en su desarrollo deberían estar amparados en algún elemento contractual que, cuanto menos, asegurara los usos previstos, de forma que en caso de que la institución deseara introducir variaciones, lo debiera hacer a través de alguna adaptación en la que la responsabilidad quedara circunscrita a quienes lo promuevan y utilicen. Este tipo de situaciones, que conlleva la progresiva desvirtuación de procesos que se diseñaron para propósitos diferentes, por desgracia es habitual y va más allá de la implicación de la validez. Este es un elemento que se ha desarrollado en aquellos contextos en los que la Evaluación presenta un mayor nivel de consolidación social y profesional, particularmente en los EE.UU. de Norteamérica.

objeto y finalidad de la evaluación, realizando una traducción *fidedigna* de las mismas —de forma que, por ejemplo, estén todas las dimensiones que aparecen en el problema y que se seleccione una información relevante para las mismas—. Y por otra parte, la *generalidad* de la traducción implica atender a la validez externa del plan de trabajo. Es decir, supone un planteamiento de las claves del mismo desde la perspectiva no sólo de la situación concreta que se aborda, sino que es conveniente atender a las soluciones observadas en situaciones de la clase a la que pertenezcan.

En la Evaluación de Programas es frecuente que se produzca una tensión entre ambos principios, *fidelidad y generalidad*, más agudizada que en otros ámbitos de trabajo. En este caso, el proceso se mueve por una demanda concreta de un «*cliente*» sobre una situación específica, así como otras audiencias. Ello conlleva que en ocasiones se mantenga la fidelidad en detrimento de la generalidad. No obstante, la clave de la validez en este caso es la fidelidad, si bien la generalidad puede constituir un componente que, de poderse dar de forma concomitante, facilitaría la explicación e interpretación de resultados al disponer de elementos de comparabilidad y contextualización externos a la propia evaluación.

En este caso, la fidelidad de la traducción del problema debe asegurarse en función de los siguientes componentes:

— Determinación de la estructura del diseño: determinación de los roles de las variables e indicadores sobre las que se sustentan, momentos de recogida de información, evaluabilidad de las informaciones, etc...
— Selección de indicadores adecuados,
— Identificación de la mejor fuente de información, en cada caso, para cada indicador.

- *Recogida de Información.*

Se trata de realizar la recogida de información atendiendo a las condiciones en que pueda asegurarse la máxima *calidad y representatividad* de la misma, en virtud de los objetivos y finalidades de la evaluación, no produciendo para ello disrupciones en la dinámica habitual del Programa. Así, en esta etapa existen diversas condiciones a tener en cuenta:

— Conseguir una relación adecuada entre la *fuente* que nos va a dar la información y el *procedimiento* que vamos a utilizar para recogerla.
— Seleccionar los instrumentos y técnicas de recogida de información de acuerdo a criterios de bondad de los mismos (fiabilidad y validez).
— Recoger la información en el momento más adecuado considerando tanto la no intromisión en el desarrollo del Programa, como la calidad derivada del mismo (por ejemplo, evaluando el impacto de un programa de inserción, momentos de recogida de información posteriores a la realización del Programa: a corto, medio y largo plazo...).

— Utilizar, en lo posible, estrategias que permitan el contraste y validación de la información, a través de fuentes, procedimientos y/o instrumentos múltiples (por ejemplo, triangulación, validación convergente...).

- *Análisis de Datos, síntesis de la información y su preparación para la Toma de Decisiones.*

Las dimensiones básicas a atender como criterios de calidad del proceso serían garantizar la *consistencia* de la síntesis de información con el planteamiento del problema y la *credibilidad* de la misma para las audiencias implicadas. En esta etapa, los componentes básicos son los siguientes:

— La *coherencia interna* de la evaluación, se basa en el desarrollo de un proceso de análisis de datos ajustado al planteamiento del problema y uso posterior de la información derivada de la evaluación, atendiendo a los diversos roles de las variables implicadas, con el fin de ampliar al máximo los elementos de comprensión y explicación de resultados.

— La *coherencia externa*, alude a la *credibilidad* de los resultados para los diferentes actores implicados. Esto supone valorar la consistencia externa para estas audiencias de la síntesis de información que se ha llevado a término. En realidad este proceso se sitúa entre esta Etapa de actuación y la siguiente, dado que una parte del mismo se lleva a cabo dentro del Proceso de Difusión de la información.

- *Gestión de la información.* Constituye el segundo núcleo de acciones desde el que abordar el problema de los criterios de calidad del proceso de evaluación. En él diferenciamos tres grandes momentos: durante el proceso de evaluación, en su finalización y después de haberse realizado la evaluación.

- *Durante el proceso de evaluación.*

Una característica básica que es conveniente tener en cuenta a lo largo de cualquier proceso de Evaluación de Programas es la flexibilidad para adaptar el mismo en virtud de las informaciones que vayan recabándose y analizándose. Es decir, la planificación se retroalimenta a través del ajuste sucesivo de todos los componentes del Plan de Evaluación durante su implementación. Esta adaptabilidad del Plan está al servicio de la fidelidad y la generalidad del mismo, así como coadyuva en el logro de la credibilidad y utilidad de la evaluación.

- *En la finalización del proceso de evaluación: Difusión de Información.*

En la Etapa de la Difusión de la información, el objetivo central es ofrecer información —al cliente y a las demás audiencias implicadas—, sobre cuales son los resultados y los criterios básicos en que puede sustentarse la interpretación de los mismos y la toma de decisiones. En este caso, las dimensiones vuelven a ser la *coherencia* en la difusión de la información y la *credibilidad* de los medios utilizados para los agentes implicados.

El criterio de *coherencia interna* supone atender a la conexión entre los Informes y lo previamente establecido de manera consensuada en las etapas anteriores del proceso. Así, será preciso aportar la información que se haya pactado al inicio del proceso con cada audiencia, atendiendo a su grado de responsabilidad e implicación en el Programa y en la Evaluación, sus intereses y posibles usos de la información. Además se deben atender condiciones de calidad en los mecanismos de difusión de la Información: utilización de un lenguaje ajustado a la audiencia a la que se dirige, estructuración de la información adecuada a las preguntas que se pretende responder, atención a la oportunidad temporal para los objetivos básicos propuestos (Tejedor, 1997; Elliott, Fischer, y Rennie, 1999). Estos mismos aspectos deberán considerarse desde la perspectiva de la *coherencia externa* para garantizar la credibilidad de los mecanismos de la difusión de la Información a las diferentes audiencias implicadas.

- *Después del proceso: valoración global del mismo y seguimiento de las consecuencias de la evaluación.*

Un elemento adicional, pero de gran importancia, lo constituyen las estrategias metaevaluativas (De la Orden y Martínez de Toda, 1999) y las de seguimiento de los procesos de implementación de las soluciones aportadas a través de la evaluación (Rodríguez Espinar, 1997, 1998). El análisis global del proceso y la comprobación de los compromisos mantenidos respecto de las consecuencias de la evaluación pueden constituir elementos clave de validación global del Plan de Evaluación.

A MODO DE CONCLUSIÓN

En primer lugar, es preciso recordar que la *Evaluación de Programas* es algo más que una aplicación de métodos a través de los que se enumeran resultados. Hay que ser conscientes de que actualmente los investigadores no son los únicos que deciden evaluar y qué métodos utilizar. Ellos intervienen en un sistema de decisiones complejo en el que también participan otros colectivos implicados en el problema objeto de evaluación, como legisladores, administradores, planificadores, profesores y alumnos, etc... (De la Orden y Mafokozi, 1999). Por ello, es necesaria una profunda reflexión en cuanto a las soluciones metodológicas y sus componentes, en un ámbito en el que la complementariedad no puede asumirse como una mera adición de aportaciones que provienen desde diversas opciones epistemológicas.

Asimismo, la atención prioritaria que debe prestarse a las características del objeto evaluado, las finalidades de la evaluación, sus usos, etc..., como elementos que deben conducir en la selección de la solución metodológica desde la que abordar el problema, debe acompañarse de una reflexión respecto al rol que desarrollamos los que, desde la Evaluación, participamos en estos procesos. Los valores, la ideología, la forma de estructurar cognitivamente la información —producto de nuestra formación metodológica y de nuestra forma de ser— inciden en los planteamientos que realiza-

mos (House, 1994). Por ello, es conveniente asumir como punto de partida un acercamiento abierto y tolerante, que no es patrimonio exclusivo de ninguna posición ideológica o epistemológica más en la práctica que en el discurso teórico en que se basa. La formación y profesionalización en el ámbito de la Evaluación de Programas puede constituir un instrumento desde el que acompañar este proceso de mejora de este tipo de procesos (Scriven, 1994; González Such, Pérez Carbonell y Perales, 2000). No obstante, el camino está por recorrer, si bien se identifican lineas de trabajo prometedoras que pueden abocar en soluciones ciertamente alternativas que superen las dificultades existentes.

BIBLIOGRAFÍA

Aliaga Abad, F. (2000). *Bases epistemológicas y Proceso de Investigación psicoeducativa.* Valencia: C.S.V.

Alvira, F. (2ª Ed.) (1996). *Metodología de la Evaluación de Programas.* C.I.S. Madrid: Cuadernos Metodológicos, 2.

Apodaka, P. (1999). Evaluación de los resultados y del impacto. *Revista de Investigación Educativa, 17,* 2, págs. 363-377.

Bericat, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social. Significado y medida.* Barcelona: Arial Sociología.

Buendía, L. (1998). La Investigación sobre Evaluación Educativa. *Revista de Investigación Educativa, 14,* 2, págs. 5-24.

Buendía, L. (2000). Ética de la Evaluación. Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas. Valencia: en prensa.

Chelimsky, E. y Shadish, W.R. (1997). *Evaluation for the 21st Century.* Sage Publications.

Colás, M.P. (1995). Diseño y evaluación de programas. En J.L. Rodríguez Diéguez y O. Sáenz Barrio, *Tecnología educativa. Nuevas tecnologías aplicadas a la educación* (pp. 45-66). Alcoy: Marfil.

Cook, T.D. y Reichardt, Ch.S. (1986). Hacia una superación del enfrentamiento entre los métodos cuantitativos y cualitativos. En: T.D. Cook y CH. S. Reichardt (1986). *Métodos Cualitativos y Cuantitativos en Investigación Evaluativa.* (3ª Ed.) Madrid: Morata.

Cook, T.D. y Shadish, W.R. (1994) Social experiments: some developments over the past fifteen years. En *Annual Review of Psychology, 45,* págs. 545-580.

Creswell, J.W. (1994). *Research Design. Qualitative & Quantitative Approaches.* Thousand Oaks. CA. Sage Publications.

Creswell, J.W. (1998). *Qualitative inquiry and research design. Choosing among five traditions.* Thousand Oaks, CA: Sage.

De la Orden, A. (1990). Evaluación de los Efectos de los Programas de Intervención. *Revista de Investigación Educativa, 8, 16,* págs. *61-76.*

De la Orden, A. y Mafokozi, J. (1999). La investigación educativa: naturaleza, funciones y ambigüedad de sus relaciones con la práctica y la política educativas. *Revista de Investigación Educativa, 17,* 1, págs. 7-29.

De la Orden, A. y Martínez de Toda, M.J. (1991) Metaevaluación educativa. *Bordón,* 43, 4, págs. 7-29.

De Miguel, M. (1988). Paradigmas de la investigación educativa española. En I. Dendaluze (Coord.), *Aspectos metodológicos de la investigación educativa* (págs. 60-77). Madrid: Narcea.

De Miguel, M. (1995). Participación ciudadana en la evaluación de las políticas públicas. *Aula Abierta, 66*, págs. 211-218.

De Miguel, M. (1999). La evaluación de programas: entre el conocimiento y el compromiso. *Revista de Investigación Educativa, 17*, 2, págs. 345-348.

Dendaluze, I. (1991). *Dimensiones e investigabilidad de la Calidad de la Educación*. En P. Braza y otros. (Eds.). Actas del I Congreso Internacional sobre Calidad de la Educación Universitaria. Cádiz: Universidad de Cádiz.

Dendaluze, I. (1995). Avances en los métodos de investigación educativa en la intervención psicopedagógica. *Revista de Investigación Educativa, 26*, págs. 9-32.

Dendaluze, I. (1997). La validez teórica de las investigaciones empíricas en Ciencias Sociales. *Revista internacional de estudios vascos, 42*, 1, págs. 77-101.

Dendaluze, I. (1998). Algunos retos metodológicos. *Revista de investigación educativa, 16*, 1, págs. 7-24.

Elliott, R., Fischer, C.T., y Rennie, D.L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology, 38*, págs. 215-229.

González Such, J., Pérez Carbonell, A. y Perales Montolío, J. (2000). Perfil profesional del evaluador. Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas. Pendiente de publicación. Valencia.

House, E.R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.

House, E.R. (1992). Tendencias en Evaluación. *Revista de Educación, 299*, págs. 43-55.

House, E.R. (1994). *Evaluación, ética y poder*. Madrid: Morata.

Jornet, J.M, Suárez, J.M., y Belloch, C. (1998). Metodología de Evaluación de Programas de Formación Profesional, Ocupacional y Continua. *Leonardo Project: Eurodialogue Evaluation*.

Krathwohl, D.R., (1987). *Social and behabioral science research: a new framework for conceptualizing implementing and evaluatig research studies*. San Francisco: Jossey Bass.

Mateo, J. (1990). La toma de decisiones en el contexto de la evaluación educativa. *Revista de Investigación Educativa, 16*, 8, págs. 95-114.

Mohr, B. (2ª Ed.) (1995). *Impact Analysis for Program Evaluation*. Newbury Park, Ca: Sage Publication.

Morse, J.M. (1997). Considering theory derived from qualitative research. En J.M. Morse (Ed.), *Completing a qualitative project. Details and dialogue* (págs. 163-189). California: Sage.

Perales, M.J. (2000). *Enfoques de Evaluación de la Formación Ocupacional y Continua. Estudio de Validación de un modelo*. Tesis Doctoral. Universitat de València. (en prensa, Microfichas).

Pérez Carbonell, A. (1998). *Metodología de la Evaluación de Programas: Evaluación de programas de postgrado en la Universitat de Valencia.Col·lecció:* Tesis Doctorals en microfitxes. Universitat de Valencia: SERVEI DE PUBLICACIONS ETD, S.A. (publicación: 1999)

350 A

Pérez Juste, R. (2000). Evaluación de programas educativos. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas*. Valencia, Marzo (en prensa).

Reichardt, C. y Rallis, S. (Eds.) (1994). *The qualitative-quantitative debate: new perspectives*. S. Francisco: Jossey Bass.

Rodríguez Espinar, S. (1997). La evaluación institucional universitaria. *Revista de Investigación Educativa, 15*, 2, págs. 179-214.

Rodríguez Espinar, S. (1998). El proceso de evaluación institucional. *Revista de Educación, 315*, págs. 45-65.

Rodríguez, G., Gil, J., y García, E. (1996). *Metodología de la investigación cualitativa*. Archidona (Málaga): Aljibe.

Rogers, P. y Owen, J. (1999). *Program evaluation. Forms and approaches*. Sage.

Rutman, L. (1984). *Evaluation Research Methods: A basic Guide*. Inc. Beverly Hills, California: Sage Publications.

Scriven, M.S. (1994). Evaluation as a discipline. *Studies in Educational Evaluation, 20*, págs. 147-166.

Suárez, J.M. (1989). *Proyecto Docente de Investigación*. Proyecto inédito, Universitat de Valéncia.

Swanborn, P.G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality and Quantity, 30*, págs. 19-35.

Tejedor, F.J. (1997). Informes en la Evaluación Institucional. *Revista Española de Pedagogía, 208*, págs. 471-498.

Tejedor, F.J., GArcía-Valcarcel, A. y Rodríguez, M.J. (1994). Perspectivas Metodológicas Actuales de la Evaluación de Programas en el Ámbito Educativo. *Revista de Investigación Educativa, 23*, págs. 93-127.

351

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 357-379

EL ANÁLISIS DE DATOS EN LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS

*Luis Lizasoain Hernández y Luis Joaristi Olariaga**
Universidad del País Vasco-Euskal Herriko Unibertsitatea

RESUMEN

En primer lugar, se aborda la caracterización del análisis de datos en el contexto de la evaluación de programas educativos. Para ello, se parte de la consideración de la evaluación de programas como investigación aplicada, de las etapas de la misma y de los tipos de evaluación.

En un segundo apartado se presentan algunas técnicas concretas como la regresión y discriminación, la segmentación, las técnicas factoriales, las series temporales, las ecuaciones estructurales y, sobre todo, los modelos jerárquicos lineales y el meta-análisis.

Por último, se discute lo relativo a la comunicación y difusión de datos y resultados distinguiendo dos audiencias: el resto de evaluadores y demás miembros de la comunidad científica, y los clientes, interesados o implicados.

ABSTRACT

First of all, the characterization of the data analysis in the context of the educational program evaluation is discussed. For this, program evaluation is considered as applied research, and the stages of the process and the evaluation types are examined.

The second part is devoted to present some techniques such as regression and discrimination, segmentation, the factorial techniques, the time series, the structural equations models and, above all, the hierarchical linear models and the meta - analysis.

Finally, questions related to the data and results communication, presentation and dissemination are discussed focusing the differences between two audiences: the evaluators, members of the scientific community, and the clients, users or stakeholders.

* Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación. Avda. de Tolosa 70. 20018 San Sebastián. E-mail: plplihel@sf.ehu.es

CARACTERIZACIÓN DEL ANÁLISIS DE DATOS

A nuestro entender, una manera adecuada de descubrir las características específicas del análisis de los datos es situarlo en el contexto general de la evaluación de programas y, a partir de ese punto, comenzar la búsqueda de las notas distintivas que lo caracterizan.

Y, por extraño que pueda parecer, hay una cuestión en la que todos los autores básicamente coinciden, y es en el hecho de considerar la evaluación de programas como una actividad investigadora.

En consecuencia, si asumimos lo antedicho, al análisis de datos le correspondería la tarea que habitualmente se cita cuando se habla de esta fase del proceso de investigación empírica en ciencias sociales y humanas. A saber: proporcionar evidencia suficiente o respuesta a las hipótesis y preguntas planteadas mediante técnicas analíticas (habitualmente estadísticas en el paradigma cuantitativo) transformando los datos en resultados.

Pero evidentemente la cuestión no es ni mucho menos tan sencilla. La evaluación de programas es una actividad investigadora, pero evaluación e investigación no son términos intercambiables. Un primer matiz es el de considerar la evaluación de programas como un caso típico de investigación aplicada hablándose también de investigación evaluativa.

La evaluación de programas como investigación aplicada

De la contraposición entre la investigación básica y la aplicada surgen bastantes de las notas distintivas que habitualmente se asignan a la evaluación de programas. Desde la perspectiva del análisis de datos, señalemos las siguientes:

En primer lugar, y como marco general de referencia, el hecho de que el objetivo primordial de la investigación aplicada no es la búsqueda de la *verdad* o la producción de conocimiento, sino el uso, **el empleo de dicho conocimiento para la acción, la intervención, en definitiva, la mejora**. Y esta última referencia a la mejora cobra especial significación en el caso de la evaluación de programas educativos (De la Orden, 1985; Tejedor, 1994; Pérez Juste, 1995, De Miguel, 1999).

De este carácter aplicado se derivan consecuencias importantes a la hora de plantearse la fase del análisis de datos. Así, las técnicas analíticas que se empleen deben estar claramente orientadas a proporcionar respuesta y evidencia a las preguntas y cuestiones que se hayan planteado.

Pero además, en el caso de las técnicas estadísticas, esta respuesta no basta con que sea estadísticamente significativa, ha de tener **significación *sustantiva*** (Botella y Barriopedro, 1995: p. 197) o *práctica* (Bickman y Rog, 1998: p. xi).

En investigación evaluativa hay que saber si los efectos tienen la magnitud suficiente como para considerarlos plenos de significado, no simplemente estadísticamente significativos.

Pero es que, además de que las técnicas analíticas nos han de proporcionar resultados *sustantivamente significativos*, en evaluación de programas hay que afrontar el hecho de que dichos resultados tienen que dar clara respuesta a las cuestiones planteadas.

O dicho en otros términos: esa significatividad sustantiva, ¿garantiza, aunque sea con un cierto margen, que el programa ha producido efectos? (Bickman y Rog, 1998: p. xiii). ¿Podemos hablar de **causalidad** en la evaluación del impacto, (Apodaca, 1999), o de atribución causal de los efectos observados (Gaviria, 1999)? ¿Dónde ponemos el énfasis, en la validez externa o en la interna?, ¿Qué grado de control es deseable o simplemente viable?

Existe acuerdo en que en la investigación aplicada, los estudios cuasi-experimentales son los más frecuentes. Desde la perspectiva de la evaluación de programas educativos, el empleo de este tipo de diseños supone muchas veces el máximo grado de control alcanzable.

En el fondo, todos estos términos, todas estas preguntas, nos remiten a la **generalizabilidad** —o en término de Cronbach, **extrapolabilidad**— de dichos resultados.

Evidentemente en todas las dicotomías que aquí planteamos subyace un continuo de forma que las respuestas a estas preguntas varían en cada caso. Pero a nadie se le escapa que todas ellas tienen importantes repercusiones a la hora de plantearse el análisis de los datos.

Pasemos a otro asunto. En la introducción de la obra de Bickman y Rog que ya hemos citado, la tercera diferencia metodológica entre la investigación básica y la aplicada es formulada bajo el epígrafe «multiple vs. single level of analysis» y en la misma podemos leer que «the applied researcher, in contrast to the basic researcher, usually needs to examine a specific problem at more than one level of analysis, not only studying the individual, but often larger groups, such as organizations or even societies». (Bickman y Rog, 1998: p. xiv).

Si esta afirmación es válida en la investigación aplicada en ciencias sociales en general, en lo tocante a la evaluación de programas educativos constituye casi un axioma.

De la Orden (1999), hablando de los modelos jerárquicos lineales afirma que «Constituyen un marco de análisis natural para el tratamiento de datos con una estructura multinivel o anidada, situación muy común dentro de la investigación educativa» (p. 425).

Más adelante tendremos ocasión de detenernos con más detalle en estos modelos. Baste ahora señalar esta **estructura jerárquica o anidada** como una de las notas distintivas más importantes a la hora de caracterizar el análisis de los datos de las evaluaciones de programas educativos.

Llegados a este punto quizá alguien esté ya planteándose si cuando hablamos del análisis de datos nos referimos exclusivamente al análisis *estadístico* de datos. Avanzamos que rotundamente la respuesta es no.

Si hay una familia semántica que configura una de las notas características de la evaluación de programas, es la que se estructura alrededor de términos como *complementariedad metodológica, enfoque multimétodo, «multiplismo», debate cualitativo-cuantitativo, triangulación, superación del debate de paradigmas*, etc.

No vamos a extendernos en este punto puesto que en este mismo número hay un excelente artículo dedicado exclusivamente a este asunto. Nos limitamos a constatar aquí que si algo caracteriza a la metodología de la evaluación de programas es esta cuestión. Prácticamente todos los autores que han dedicado unas líneas a la evaluación de programas coinciden en señalar **la complementariedad como enfoque más**

acertado de forma que el análisis de los datos debe llevarse a cabo empleando de forma complementaria o simultánea **múltiples enfoques, métodos y técnicas**. Con respecto a los paradigmas, inmejorable resumen lo constituyen las normas elaboradas por el *Joint Committee* (1994) cuando al tratar del análisis de la información, ofrecen dos normas de idéntico texto con la única diferencia de que la A8 se refiere a la información cuantitativa y la A9 a la cualitativa.

Somos conscientes de que el asunto tiene en cualquier caso sus complicaciones: para algunos, el hablar de *múltiples enfoques* supone que es deseable o incluso necesaria la complementación mutua entre el enfoque cuantitativo y el cualitativo. Otros con ese término o con el de triangulación hacen referencia al uso de variedad de técnicas *dentro de un solo paradigma*. Es, por ejemplo, el sentido que Cook otorga al término «*multiplismo*» y que como tal se adopta a lo largo de la obra de Fernández Ballesteros (1995) asumiendo un enfoque cuantitativo.

Como muestra del uso de técnicas tanto cualitativas como cuantitativas a la hora de analizar los datos valgan los siguientes ejemplos:

Freixa y otros (1996) emplean técnicas estadísticas para datos textuales al realizar el análisis de contenido de las respuestas a entrevistas semi-estructuradas. Sogunro (1997) combina técnicas estadísticas clásicas con el análisis de contenido de los datos cualitativos. Yin (1998) al tratar del estudio de casos expone la posibilidad de emplear unidades de análisis anidadas como muestra de la factibilidad de simultanear un estudio cualitativo y cuantitativo. Fetterman (1998) al hablar de los análisis propios de los estudios etnográficos (pp. 494-498) cita entre otras la triangulación, elaboración y comparación de patrones, los sucesos-clave, el uso de mapas y gráficos, el análisis de contenido, la cristalización y la *estadística* especialmente las técnicas no paramétricas. Otro ejemplo es el artículo de Knight y Kuleck (1999) titulado *The value of multimethod qualitative/quantitative research methodology in an educational program evaluation: a case study.* Hemos transcrito el título completo porque pensamos es lo suficientemente claro como para requerir comentario adicional alguno. Por último, González Ramírez (2000) realiza análisis estadísticos (descripciones univariadas y análisis de varianza) sobre los resultados de los alumnos y análisis de contenido para evaluar los resultados desde la perspectiva de los profesores.

Sin negar la complejidad que el asunto tiene, con estos pocos ejemplos hemos querido simplemente mostrar que el mestizaje (usando un término muy en boga) metodológico es una de las características más específicas del análisis de los datos en investigaciones evaluativas.

La última nota distintiva que nace de la contraposición entre investigación básica/investigación aplicada, se refiere a los agentes que intervienen en ésta.

Simplificando un tanto las cosas, en un esquema investigador tradicional habitualmente se distingue sólo entre un agente (el investigador o equipo investigador) y un paciente (los sujetos o unidades investigadas). (Ciertamente no hay que olvidarse de los que evalúan y financian —o no— las investigaciones, pero éste es otro asunto). Aquí el proceso se desarrolla bajo el control del investigador y los resultados se comunican a otros posibles agentes mediante la difusión por los canales habituales de comunicación de la comunidad científica formada por dichos agentes.

355

Por el contrario, en la investigación aplicada, y más en la evaluación de programas, es extremadamente frecuente que a este esquema haya que añadir un elemento más. Los términos varían mucho, pero los más frecuentes son: *clientes, audiencias, interesados, implicados, participantes,* etc.

La distinción terminológica es importante porque trasluce muy diferentes enfoques en la forma en que se aborda el proceso evaluativo y el papel asignado a los diferentes participantes en el mismo.

No es lo mismo el típico «encargo» que un evaluador profesional realiza como evaluador externo, que un proceso evaluador participativo en el que el hipotético papel del evaluador sea el de facilitador, animador o coordinador de un grupo de implicados que son parte activa y los verdaderos agentes de la, en este caso, autoevaluación.

Entre estos dos extremos se dan muchas situaciones intermedias en las que es frecuente que los interesados participen en algunas fases del proceso de evaluación, y en el caso de las evaluaciones educativas, el énfasis puesto en la mejora recomienda optar por evaluaciones internas realizadas por los mismos profesores (Pérez Juste, 1995: p. 112). Pero no es menos cierto que frecuentemente se realizan evaluaciones externas y que, sobre todo, las condiciones muchas veces aconsejan un enfoque mixto.

De una manera o de otra, independientemente de lo participativo que se pueda considerar el enfoque adoptado, lo importante es hacer notar aquí que tanto alguna de las fases del proceso evaluador como los resultados del mismo han de ser comunicados y compartidos con el conjunto de **agentes *implicados***.

En el caso concreto de los resultados, esto supone que han de ser transmitidos a dos audiencias muy distintas: al resto de la comunidad científica, pues como investigación que es debe ser difundida siguiendo los canales y normas al uso; y a los *clientes, implicados* o *interesados,* y aquí el asunto cambia porque este segundo tipo de audiencia no tiene por qué compartir (de hecho normalmente no lo hace) ni el lenguaje ni el soporte técnico que subyace al proceso evaluativo.

Y el que esta comunicación sea eficaz es crucial para la evaluación de programas, pues no se trata de una cuestión accesoria nacida de un deseo o prurito de difundir o comunicar al máximo los resultados. Por el contrario, saber comunicar a los interesados los resultados del análisis es objetivo fundamental si queremos que la evaluación produzca efectos, sea de utilidad para la acción y los procesos de mejora.

Pensamos que las hasta aquí vistas son —desde la perspectiva del análisis de datos— las notas distintivas que emanan de la consideración de la evaluación de programas como investigación aplicada. Ahora, para completar este intento de caracterización, centremos nuestra atención en el proceso evaluativo en sí: en sus etapas y fases y en los tipos de evaluación que de las mismas se derivan.

Las etapas del proceso y los tipos de evaluación

Dado que la evaluación de programas es una actividad investigadora, la consecuencia que de ello se deriva es que las fases de la misma no han de diferir excesivamente de las propias del proceso investigador.

Y de hecho así ocurre. Si examinamos en la bibliografía esta cuestión, o sin ir muy lejos, si observamos el índice de este número monográfico, nos daremos cuenta que casi siempre, después de las decisiones sobre el diseño y de la etapa de recogida de información, aparece la fase del análisis de los datos, a la que suele seguir la elaboración del informe y la puesta en práctica de las conclusiones.

Y, por descontado, nada hay que objetar a esta secuencia en la medida en que refleja el orden lógico del proceso investigador. Los matices suelen surgir en el número de etapas que se distinguen, o en las diferencias según los enfoques de la evaluación.

Pero desde nuestra particular perspectiva centrada en la fase del análisis de los datos, nos interesa señalar que, en primer lugar, las decisiones sobre qué técnicas emplear para analizar los datos van a estar a su vez, lógicamente en función de las decisiones que en las fases previas se hayan ido tomando: qué enfoque se ha dado a la evaluación (cuantitativo/cualitativo/mixto), qué preguntas o cuestiones se plantean, qué diseños se han empleado, qué unidades se han definido, cómo han sido seleccionadas y —si ha sido el caso— asignadas, qué tipo de información y mediante qué instrumentos se ha recogido.

Segundo, si bien esta secuencia es lógicamente consistente, no debe ser tomada al pie de la letra en el sentido de que después de recoger la información, ésta se organiza y se analiza dando respuesta a las preguntas planteadas. Evidentemente el proceso no es ni mucho menos tan lineal porque el análisis de los datos es un **proceso abierto, cíclico e interactivo**: los resultados de los análisis preliminares condicionan los siguientes y así sucesivamente, a menudo modificando las previsiones iniciales.

Este carácter abierto de la fase del análisis es válido y aplicable a cualquier proceso de investigación, aunque lógicamente en mayor grado cuanto más exploratorio sea el enfoque adoptado. Pero en el caso de la evaluación de programas cobra especial relevancia habida cuenta del papel que los agentes implicados o interesados pueden jugar en esta fase. Cuanto más abierto y participativo sea el enfoque, más interactivo y cíclico es el análisis de los datos.

Pero además de esta ordenación lógica basada en las fases del proceso investigador, también es posible distinguir etapas en función del desarrollo, de la secuencia del propio programa a evaluar. Aquí, en función de la fase del proceso en que se centre la actividad evaluadora, nos podemos encontrar con distintos tipos de evaluación: evaluación de necesidades; evaluación del diseño; evaluación de la evaluabilidad; evaluación de la implementación; evaluación de la cobertura; evaluación del proceso; evaluación de resultados, impactos o producto; y, evaluación económica.

Desde nuestra habitual perspectiva centrada en el análisis de los datos, en primer lugar señalemos que el análisis de datos encuentra su lugar, digamos, natural en la fase o contexto de la **evaluación de resultados** o de impactos.

Segundo, que al igual que antes, las técnicas analíticas a emplear estarán también en función de los resultados y de las consiguientes decisiones que se hayan ido tomando en evaluaciones previamente realizadas (del proceso, por ejemplo).

Tercero, que las técnicas analíticas pueden emplearse —y de hecho se emplean— en muchas fases o tipos de evaluación, *y no sólo en la evaluación de resultados o de impactos* a pesar de que algunos consideren ésta la fase analítica por excelencia. Lógicamen-

te también, las técnicas analíticas a emplear estarán en función del tipo de evaluación en cuyo contexto se lleven a cabo. Por ejemplo, Sánchez y Anguera (1993) analizan las posibilidades que la técnica del PERT brinda en la fase de la implantación y planificación de la evaluación de programas.

En el ámbito de la evaluación económica, ejemplos recientes de lo dicho encontramos en Francis y otros (1999) que llevan a cabo un análisis de costo-eficacia de un programa de tratamiento de drogadicciones, o en Plotnick y Deppman (1999) que abordan los análisis de costo-beneficio en evaluación de programas en general y en programas de prevención del maltrato infantil en particular.

En lo relativo al análisis de datos en la evaluación de necesidades, destaca aquí lo referente al manejo y análisis de datos secundarios. A saber: indicadores, estadísticas oficiales, etc.

Se trata de instrumentos fundamentales para la evaluación y tanto si son empleados en el análisis de necesidades como en la estimación de los impactos de un programa, el evaluador puede optar por elaborar un sistema de indicadores *ad hoc* o por emplear algunos sistemas de indicadores ya elaborados por diversidad de organismos e instituciones. Lázaro (1992) analiza las condiciones que los mismos deben reunir, compara los de diferentes organismos y examina el caso específico de los indicadores en educación. Cornejo y Martínez (1992) estudian las posibilidades que brinda el manejo sistemático de fuentes secundarias de información en evaluación de programas de intervención social y ofrecen una muy interesante aplicación del análisis de correspondencias a los datos del censo de población.

Por último, con respecto a la evaluación de resultados, no podemos por menos que asumir que éste sigue siendo el tipo de evaluación o la fase del proceso evaluativo en que se cumple el objetivo clásico del análisis de los datos: verificar, con un grado de confianza aceptable, que el programa permite obtener los resultados planteados.

Todas las cuestiones que anteriormente hemos apuntado relativas a la causalidad, la validez interna, el control, los diseños, la integración metodológica, etc. encuentran aquí perfecto acomodo.

Botella y Barriopedro (1995) apuntan cuatro funciones principales del análisis de datos en la valoración del impacto: descripción comprensiva de datos brutos; valorar si las relaciones detectadas son o no significativas; estimación de magnitudes y diferencias; y, estimar la sustantividad de los efectos.

Desde esta perspectiva, muchas son las técnicas analíticas que aquí se pueden emplear y, centrándonos en las técnicas estadísticas, más adelante trataremos algunas de ellas. Pero no queremos dejar pasar este punto sin justificar el por qué en dicha exposición no vamos a tratar de los análisis descriptivos básicos y de las técnicas inferenciales —paramétricas y no paramétricas— clásicas.

La razón es evidente: creemos que son cuestiones suficientemente conocidas y perfectamente desarrolladas por prestigiosos autores en obras y manuales ampliamente difundidos.

Y ésta es la única justificación. Queremos dejar claro que su ausencia no implica que estén en desuso (muy al contrario como ahora veremos) por ser inadecuadas, excesivamente sencillas o simplemente pasadas de moda. Asumimos además que

estas dos últimas pretendidas *razones* no son tales, pero, lamentablemente, pensamos que hay que incluirlas pues todavía a estas alturas hay quien se rige por criterios de moda o actualidad o quien piensa que la sencillez es un defecto.

Como muestra de su uso y adecuación, en las siguientes referencias encontramos estudios evaluativos con diseños pretest-postest en que el análisis de los datos se efectúa mediante pruebas paramétricas y no paramétricas de contraste de medias: Hernández, Aciego y Domínguez (1994), Echeburua y otros (1996), Herráiz (1998), Pérez y Salvador (1998), y Filella (2000).

ALGUNAS TÉCNICAS ESTADÍSTICAS PARA EL ANÁLISIS DE DATOS

Hecha esta relación de notas distintivas y de características deseables del análisis de datos en la evaluación de programas educativos, a continuación vamos a comentar someramente algunas técnicas y procedimientos de análisis que pensamos reciben o deben recibir un tratamiento especial.

Se trata, como no podía ser de otra manera, de una selección personal responsabilidad de quienes esto firmamos, y, por tanto, sesgada. En algunos casos hemos incluido técnicas o procedimientos que consideramos de presencia inexcusable como lo relativo al meta-análisis o los modelos jerárquicos lineales. Muy probablemente todo el mundo podrá coincidir en ello. Quizá no ocurra lo mismo en otros casos en que su aparición obedece a razones más parciales: nuestra personal opinión o nuestra experiencia investigadora.

Como ha quedado dicho, no vamos a ser exhaustivos; lo que a continuación vamos a hacer es exponer aquellas particularidades de las técnicas más generales en el sentido de su aplicabilidad a problemas de muy diversos tipos.

No se puede decir que hay técnicas específicas de análisis de datos para la evaluación de programas (Tejedor, 1994). Lo que sí es cierto es que en función del tipo de evaluación algunas son más apropiadas que otras, de manera que en algunos son más convenientes técnicas asociadas a diseños transversales y en otros, longitudinales.

Así, si en algún caso se requiere de «jueces» son útiles técnicas como el escalamiento multidimensional; o el análisis de datos textuales para el tratamiento estadístico de cuestiones abiertas, suponiendo que se pretende superar lo simple con un contenido exploratorio considerable. Si, por otra parte, algún aspecto de la evaluación se desarrolla a lo largo del tiempo, la diversidad de técnicas cuantitativas es amplia, empezando desde las técnicas asociadas a los diseños experimentales y cuasi-experimentales, con sus pruebas paramétricas y no paramétricas de comparación de dos o más grupos, pasando por modelos de medidas repetidas, series temporales, tablas múltiples, etc.

A la postre, el diseño de la evaluación también depende de las herramientas de que se disponga para el análisis. Es cierto que quien lo desconoce todo sobre la metodología cuantitativa suele, por lo común, plantear diseños cuyos datos no hay quien analice. Pero también ocurre lo contrario: a veces se recurre a diseños muy simples en los que se pierde voluntariamente la complejidad de la realidad.

Nuestra idea no es desarrollar los distintos tipos de técnicas sino, con el fin de ser útiles, centrarnos en los que son más generales. Otros criterios que tenemos presentes

359

son la sencillez de interpretación, así como que tengan condiciones de aplicación poco restrictivas, pues ya sabemos lo frustrante y cansado que es el análisis de los residuos.

Análisis discriminante, regresión logística. Regresión y discriminación no paramétricas (técnicas de segmentación)

Desde una perspectiva paramétrica, si la variable dependiente es dicotómica y las independientes son cuantitativas, podremos aplicar las dos primeras técnicas citadas. Si se plantean dudas acerca de si aplicar una o la otra, no está de más recordar que siempre es más útil disponer de diversos puntos de vista por lo que lo recomendable es realizar ambos y comparar sus resultados. Pero en tal caso conviene tener presente que el análisis discriminante precisa de condiciones de aplicación ante las que la regresión logística es más robusta (en ésta sólo hay que tener precaución con la colinealidad). Un estudio detallado de ambas técnicas se puede encontrar en Hair, Anderson y otros (1999). Hannan y Murray (1996) llevan a cabo una simulación con métodos Monte Carlo comparando el rendimiento de los análisis lineales y logísticos.

Se trate de regresión o de análisis discriminante, lo habitual es utilizar sus versiones paramétricas, tanto si se trata de explicar y predecir como si el fin es explorar. Cuando los datos de una evaluación son no experimentales es preciso explorar. Y en este caso se dispone de un conjunto de técnicas, libres de restricciones y, aunque con un proceso de computación tedioso, de una gran sencillez de interpretación. Nos estamos refiriendo a la segmentación. La variable dependiente (única, pues no hay generalización multivariante) puede ser de cualquier naturaleza así como las independientes o predictoras. Consiste, en pocas palabras, en dividir progresivamente la muestra en clases, pero está más relacionada con la regresión y la discriminación que con las técnicas de clasificación, pues se trata de encontrar los grupos de casos que mejor expliquen las modalidades de una variable cualitativa o los valores de una cuantitativa.

Las técnicas de segmentación son muy usadas en estudios de mercado. En nuestro campo, a pesar de las ventajas que hemos apuntado, no lo son tanto. Una excepción es el trabajo de Repetto y otros (1994) donde emplearon el modelo CHAID en el análisis de los datos de la evaluación de un programa de mejora de la comprensión lectora.

El análisis factorial y las tablas múltiples

La segmentación es una forma de explorar considerando las variables dependientes de forma aislada. De alguna manera se pierde la multidimensionalidad de la realidad. Por ello convendría recurrir a los modelos exploratorios por excelencia. Esto es, el análisis factorial (análisis factorial en factores comunes y específicos así como componentes principales y análisis factorial de correspondencias).

Dada la naturaleza de las variables a considerar en todo fenómeno social o educativo y siguiendo a los ya citados Cornejo y Martínez (1992), el análisis de correspondencias es una eficaz herramienta para la descripción y evolución en el tiempo de los fenómenos macrosociales implicados en la evaluación de programas. Para ilustrar dicha afirmación, aplican el análisis de correspondencias a los datos del censo de pobla-

ción operando con once índices de riesgo de desadaptación social en la infancia extraídos de dichas fuentes con el fin de caracterizar las situaciones de riesgo potencial.

Otros estudios evaluativos en que se emplean estas técnicas son los realizados por Carballo (1990), y por los también previamente citados Freixa y otros (1996); y Kreft (1998) que usa la terminología sajona refiriéndose al análisis de correspondencias como análisis de homogeneidad.

Pero además de la perspectiva apuntada, hay que tener en cuenta que en alguna etapa de una evaluación nos vamos a encontrar con varias tablas de datos, dado que va a haber un antes y un después; eso cuando menos, pues nos podemos encontrar con una serie temporal. Cada tabla o grupo de variables puede corresponder a instantes o contextos diferentes (Lebart, Morineau, Piron, 1995).

Según los análisis factoriales clásicos, en que todas las tablas se analizarían conjuntamente o de forma yuxtapuesta, no se podrían estudiar globalmente las estructuras, tanto comunes como específicas de estos grupos de variables, ni construir tipologías de los individuos ni de las variables. Bajo este enfoque multivariante asociado al análisis factorial, sugerimos que la compleja realidad evolucionando a través del tiempo será más adecuadamente analizada por medio del análisis de tablas múltiples. Se analiza un conjunto de individuos caracterizado por varios grupos de variables (por ejemplo el pretest y el postest tratados multidimensionalmente; bajo el análisis factorial de las tablas yuxtapuestas puede ocurrir que las tipologías de los individuos estén construidas sólo según unas cuantas variables del pretest y, sin embargo, nos interesa analizar cómo evolucionan los individuos a través del tiempo, caracterizándolos por variables de ambos grupos). Al no tratar las tablas de forma exclusivamente conjunta, permite representar los individuos según cada grupo de variables, descubrir qué tablas (o grupos de variables) se asemejan, proporcionar una tabla que sintetice el conjunto y describir la dispersión entre tablas con relación a la sintética y determinar el origen de tales desviaciones. Para profundizar en estas técnicas cabe mencionar la obra de Escofier y Pagès (1990).

Pero además de estas técnicas, hay modelos creados específicamente para estudiar la evolución de un fenómeno a lo largo del tiempo.

Diseños de medidas repetidas y series temporales

Se plantean como las alternativas más intuitivas, dado que un programa se desarrolla en el tiempo. En general el desarrollo en el tiempo se basa en un número reducido de momentos.

En los diseños de medidas repetidas, analizados por medio del análisis de la varianza multivariante (Ximénez y San Martín 2000), debido a que a cada sujeto se le aplican todos los tratamientos, no se precisa del control ni de la aleatorización completa del análisis de la varianza, pero al estar inmersos en el modelo lineal general, están sometidos a las típicas condiciones de aplicación, además de que resultan «penosos» si se quiere considerar el contexto y los condicionamientos prácticos.

Por su parte una serie temporal consiste en un conjunto de observaciones de una variable en intervalos regulares de tiempo (Uriel 1992). La problemática de las series

361

temporales empieza porque se precisa una medición a intervalos regulares de tiempo, que como poco debe ser de 50; así, en los modelos ARIMA se precisa de numerosos instantes de observación, pues en caso contrario hay dificultades en la identificación del modelo. Además, las variables sólo pueden ser continuas. Las series temporales requieren procesos estabilizados y con intervenciones muy limitadas. En nuestro entorno no está de más recordar que la educación no sólo no es un proceso estabilizado, sino que no debe serlo (Sáenz, González, 1990). Pero, al margen de estas consideraciones y limitaciones, muestra del interés que ofrecen son los trabajos de Arnau, Ruiz y Salvador (1988) que examinan la evaluación del tratamiento en diseños conductuales mediante estas técnicas; o el propio Alvira (1991) que dedica un apéndice a discutir la aplicabilidad de estos modelos. Al hablar de la evaluación económica, citamos el estudio de Plotnick y Deppman (1999). En el mismo llevan a cabo un análisis de costo-beneficio de programas de prevención del maltrato infantil y proponen, entre otras, técnicas analíticas basadas en series temporales para la extrapolación de los posibles beneficios de un programa.

De cualquier forma, a la vista de la problemática de estas dos técnicas, proponemos como salidas a este callejón dos modelos que entre sus especificidades también dan cabida a la multidimensionalidad y al tratamiento de diseños longitudinales.

Modelos de ecuaciones estructurales

Según Batista y Coenders (2000), ya que los fenómenos de interés son complejos, tienen muchos aspectos, obedecen a múltiples causas y están frecuentemente medidos con error, identificar el origen de su variabilidad requiere servirse de métodos multivariantes adecuados, como los modelos de ecuaciones estructurales, que permiten incorporar el error de medida y considerar relaciones recíprocas entre constructos.

Al contrario del análisis de la varianza con medidas repetidas, en que sólo es viable una única relación entre variables dependientes e independientes, los modelos de ecuaciones estructurales posibilitan especificar las variables independientes que predicen cada una de las variables dependientes. Operando con datos no experimentales y relaciones lineales se puede entender que surgen como alternativa al modelo lineal general permitiendo además una visión holista.

Fundamentalmente se centra en la especificación del modelo a voluntad y su modificación para mejorar el ajuste a los datos (esto quiere decir confirmatorio: especificar el modelo antes de su estimación); y ello basado en la descomposición de varianzas y de covarianzas. Otro aspecto básico es el tratamiento del efecto de los errores de medida de las relaciones entre las variables, para lo cual hay que basarse en medidas repetidas. Es decir, se pueden contrastar relaciones causales entre variables, así como evaluar la calidad de la medida, en lo que se entiende como modelo de análisis factorial confirmatorio, como paso previo a las especificaciones del modelo.

Una aplicación de este tipo de modelos (en concreto del de crecimiento latente) nos lo ofrecen Wang y otros (1999) evaluando programas de intervención de reducción de conductas de riesgo con respecto al SIDA.

Los modelos jerárquicos lineales

De la Orden (1999), al introducir los trabajos sobre esta cuestión presentados al último congreso de AIDIPE, afirma que «No resulta excesivamente arriesgado afirmar que el principal avance metodológico de los últimos años en investigación en Ciencias Sociales son los Modelos Jerárquicos Lineales» (1999: p. 425).

Ciertamente, constituye una enorme ventaja el poder disponer de un conjunto de técnicas que permiten analizar los datos tomando en consideración la estructura multinivel o anidada de los mismos, habida cuenta que en el ámbito educativo ésta suele ser su conformación habitual, casi diríamos natural.

Este tipo de modelos resulta de aplicación más general, aunque en su aspecto más sencillo no cabe la multidimensionalidad. Se trata de explicar y/o predecir una variable dependiente o respuesta, que en el caso más sencillo es normal. Los individuos están agrupados según uno o más criterios o niveles entre los que se establece, habitualmente de forma natural, una jerarquía. Dado que los sujetos dentro de jerarquías tienden a ser más homogéneos que los obtenidos aleatoriamente de la población, conviene que cada nivel tenga asociado su propio modelo, que representa las relaciones estructurales así como la variabilidad residual dentro del nivel (Bryk, Raudenbush, 1992). Dicho de otra manera, la noción central es que modelos lineales diferentes para el primer nivel o micro-nivel deben ajustarse para cada contexto (macro-nivel) (De la Orden, 1999). Los coeficientes de regresión del primer nivel se analizan como variables aleatorias en el segundo nivel, descomponiéndose la varianza en la de ambos niveles. Se obtienen estimaciones estadísticamente eficientes, y desde el punto de vista práctico, se pueden ordenar los elementos del segundo nivel y superiores en función de la variable respuesta. Para profundizar sobre estos modelos remitimos a Goldstein (1995).

Con los modelos jerárquicos lineales se resuelve el problema de la unidad de análisis, puesto que se incluyen distintos niveles jerarquizados. Las variables asociadas a los distintos niveles jerarquizados afectan a los resultados individuales. Si la unidad de análisis es el individuo, se asignan las características del nivel correspondiente a cada sujeto de él, dando origen a la no independencia y a eludir los efectos de grupo. Al contrario, si se considera que la unidad de análisis es un determinado nivel, se incluyen en él las características medias de los niveles inferiores, perdiéndose así la variabilidad entre esos niveles inferiores. Además, los resultados del modelo lineal general dentro de cada elemento del segundo nivel, por ejemplo, no son generalizables más que si tales elementos constituyen toda la población.

Entre los distintos aspectos ventajosos que el modelo jerárquico lineal presenta cabe mencionar que precisa de menos condiciones de aplicación en comparación con los modelos clásicos. Además proporciona modelos más intuitivos, pues refleja el proceso de muestreo, que a menudo es por conglomerados. Es una alternativa a la técnica clásica del análisis de la varianza con medidas repetidas, que precisa de las mismas mediciones para cada sujeto y un cierto balance en la estructura; los sujetos se incluyen en el segundo nivel y las distintas mediciones en el primero, esto es, se da un anidamiento dentro de los sujetos; los resultados obtenidos con el modelo jerár-

quico lineal son más sencillos de interpretar y de comunicar. Se pueden tratar datos de respuesta discreta y variables de respuesta cuantitativas y cualitativas simultáneamente.

Si de cada individuo se miden diversas variables de respuesta (modelos multivariantes), se sitúa al individuo en le segundo nivel y las variables de respuesta en el primero; se da un correcto tratamiento a los datos ausentes, que en muchas ocasiones se deben al propio diseño, pues hay individuos en que sólo se puede medir un subgrupo de variables respuesta; los resultados son fácilmente interpretables, en comparación con los de los análisis multivariantes basados en el modelo lineal general multivariante.

Con todo lo dicho, no es de extrañar que aquí sí sea fácil encontrar abundantes referencias. Los interesados encontrarán especialmente útil la página Web que sobre el proyecto multinivel tiene el Instituto de Educación de Londres. Su dirección es *http://www.ioe.ac.uk/multilevel/* y en la misma encontrarán abundante información bibliográfica, información sobre programas, ficheros de datos, ejemplos, etc.

Pero junto a esto, tanto aplicaciones de estos modelos como estudios de índole más metodológica podemos encontrar en los siguientes artículos:

El ya citado trabajo de Kreft (1998) que efectúa un re-análisis de investigaciones evaluativas sobre prevención del consumo de drogas usando, además del ya citado análisis de correspondencias, los modelos multinivel. Krull y MacKinnon (1999) realizan un estudio de simulación centrado en las variables mediadoras. Gaviria (1999) analiza las posibilidades del enfoque multinivel en la evaluación de sistemas educativos y Murillo (1999) hace lo propio en el ámbito de los estudios sobre eficacia escolar.

Además, cada vez es más frecuente encontrar aplicaciones de estos modelos al meta-análisis: McNeal y Hansen (1995) realizan un estudio meta-analítico de las diferentes evaluaciones de los programas DARE usando técnicas multinivel. Desde un punto de vista más metodológico, en Castro y De la Orden (1998) y en Castro (1999) se analizan y comparan los diferentes modelos de meta-análisis y se estudian las aportaciones que a los mismos pueden prestar los modelos jerárquicos lineales.

Es habitual encontrar evaluaciones que consisten en distintas réplicas de un programa, ya que han sido implementados en diversos lugares y/o a distintos grupos, así como en distintos períodos de tiempo. Por otra parte, también es habitual aplicar un programa a grupos reducidos, lo que origina la práctica imposibilidad de rechazar la hipótesis nula. Todo esto es suficiente para darle al meta-análisis un lugar específico en el análisis de datos en evaluación.

El meta-análisis

El meta-análisis (Hedges y Olkin, 1985) es un conjunto de técnicas estadísticas de integración de los resultados de distintas investigaciones sobre un mismo problema. Se trata de buscar la consistencia de los mismos, de examinar las posibles contradicciones, de buscar patrones comunes, o, sin más, de resumir sus resultados cuantitativos. Para ello se parte de utilizar una métrica común entre ellos. Más técnicamente es el

361

análisis de un conjunto de estimaciones independientes con distribución normal y varianza conocida.

El proceso es análogo al de una investigación primaria. Se definen los objetivos y se clarifican los criterios de selección de estudios y datos. Tras codificar la estructura de las investigaciones en función de los objetivos de la revisión y para evitar la ausencia de validez, se unifica, como se ha mencionado, bajo una métrica común, habitualmente el tamaño del efecto.

De entre los distintos enfoques, no excluyentes, destacamos que en el basado en el *Escrutinio* el problema es que se confunden el efecto de la intervención y el del tamaño de la muestra, además de que no es considerado el impacto del programa evaluado. En el *Clásico o de Glass* (Glass, McGaw y Smith, 1981) se caracteriza cada resultado (como unidad de análisis) de cada estudio por el tamaño del efecto. La crítica se basa en que se da más peso a los estudios con más resultados, los datos no son independientes, pues cada estudio puede traducirse en distintas unidades de análisis y la fiabilidad de los resultados está condicionada por la inclusión de estudios metodológicamente defectuosos. En lo que se refiere a la conocida como efecto del estudio, que incluye criterios de selección de estudios, éstos quedan caracterizados por el tamaño del efecto; como aspectos negativos cabe mencionar los sesgos en la inclusión de estudios y el reducido número de unidades de análisis. Otro enfoque es el de las *Pruebas de homogeneidad*, desarrolladas para determinar si la varianza entre tamaños del efecto es aleatoria. El *meta-análisis Psicométrico* (Hunter y Schmidt, 1990), según el cual se recogen todos los estudios independientemente de su calidad y se corrige la distribución de los tamaños del efecto según una serie de factores que influyen sobre ella, presenta la limitación de que de cada estudio se precisa la suficiente información, no siempre disponible, para proceder a la corrección mencionada. Por último en el *Modelo jerárquico lineal*, como enfoque más novedoso, se considera que la organización de los datos meta-analíticos responde a una estructura jerárquica o anidada. (Castro, 1999) (el anidamiento se produce dentro de los estudios).

Por último, mencionemos las dificultades genéricas en el meta-análisis. Así, cabe indicar la relativa a la selección de estudios, en que el debate se establece en cuanto a si sólo incluir solamente los publicados o en función de su calidad; por otra parte surge el problema de los estudios con resultados ausentes en que las alternativas son eliminarlos totalmente contra proceder a imputación. También la presencia de múltiples tamaños del efecto en un mismo estudio, lo que da lugar a la no independencia, es resuelto de diversas formas, contradictorias entre ellas. En lo que se refiere a la magnitud del efecto, la discusión es la de siempre: a partir de qué valor un número es pequeño y a partir de cuál grande; se barajan distintas alternativas que se pueden encontrar en Cooper (1981), Cohen (1988) y Lipsey (1990).

Tal y como apuntábamos al inicio de este apartado, somos conscientes de que hemos dejado al margen otros modelos y técnicas de interés, como el escalamiento multidimensional, el logarítmico lineal, el análisis conjunto, etc.; pero creemos que con los que hemos presentado, bien exponiéndolos, bien señalando sus limitaciones, el objetivo inicialmente marcado queda, a nuestro juicio, suficientemente cumplido.

DIFUSIÓN Y COMUNICACIÓN DE LOS RESULTADOS DE LA EVALUACIÓN DE PROGRAMAS

Si repasamos algunas de las notas distintivas con que hemos caracterizado a la evaluación de programas educativos, e incluso si examinamos algunas de las técnicas analíticas que hemos presentado (como lo tocante al meta-análisis), podemos darnos cuenta que las cuestiones relativas a la difusión y comunicación de los datos y resultados de una investigación evaluativa constituyen un aspecto crucial del proceso.

Aquí hemos introducido dos términos a los que, a pesar de su evidente vecindad semántica, asignamos significados muy distintos. Se trata de los de *difusión* y *comunicación*. La diferenciación básica que establecemos es la siguiente:

Por *difusión* entendemos la presentación del proceso, los datos y resultados de la evaluación al resto de colegas, de miembros de la comunidad científica. Esta *diseminación* (traducción literal del término inglés) es llevada a cabo fundamentalmente a través de las revistas científicas especializadas.

Por *comunicación* entendemos la presentación de los resultados de la evaluación, y de cualquier otra información que se considere relevante, a los clientes, interesados, implicados, etc. Esta presentación suele efectuarse de forma oral en una o varias sesiones e igualmente suele elaborarse un informe escrito.

Queda pues claro que la distinción básica que establecemos entre estos dos términos se basa en los destinatarios o audiencias de la información. La asignación de dichos términos es poco menos que arbitraria aunque *difusión* o *diseminación* se suelen asociar a la literatura científica, de ahí su elección por nuestra parte. En cualquier caso, el significante elegido es lo de menos. Lo importante son las consecuencias que se derivan de tener que dirigirse a dos audiencias tan distintas como las que acabamos de apuntar. Y en ellas vamos a centrarnos ahora.

Difusión de los resultados

Tal y como en distintas ocasiones hemos afirmado, la evaluación de programas puede ser considerada como un caso o tipo específico de investigación aplicada. Como tal, sus resultados han de ser difundidos entre los miembros de la comunidad científica.

Y esta labor de difusión e intercambio se realiza mediante los canales habituales: de una parte las reuniones científicas, seminarios, congresos, etc., y de otra, las publicaciones en revistas especializadas. El contenido y formato de las mismas está suficientemente tratado en los manuales de investigación e incluso los editores o las sociedades científicas y profesionales establecen y publican normas al respecto.

Poco más tendríamos que decir al respecto si no fuese por una cuestión: como ya hemos tenido ocasión de comentar, en evaluación de programas la acumulación y contrastación de evidencia, o en general las labores de re-análisis, meta-análisis o metaevaluación, tienen una importancia especial. Y como ya hemos apuntado también, para llevar a cabo este tipo de análisis de segundo orden a veces no basta con la información que suele aparecer publicada.

Dicho en otros términos, en evaluación de programas —como en otros muchos ámbitos de la investigación en ciencias sociales y humanas— hay ocasiones en que, además de la información contenida en los informes de investigación en sus diferentes formatos, *es preciso o deseable disponer de acceso a los datos*.

De la Orden y Martínez de Toda (1992) insisten en esta cuestión al hablar del análisis secundario como forma más completa de metaevaluación y añaden que la disponibilidad de los datos para el re-análisis es un indicador de credibilidad del evaluador (p. 519).

Pues bien, esta disponibilidad de los datos plantea cuestiones como las siguientes: el soporte o canal mediante el cuál los datos son accesibles, el formato y estructura de los mismos, y, por último, las garantías de confidencialidad y la ética de la investigación.

Con respecto al soporte parece claro que actualmente las redes informáticas constituyen el medio más idóneo para garantizar la accesibilidad de los datos a otros investigadores. De hecho cada vez es más frecuente encontrarnos en Internet con ficheros y conjuntos de datos disponibles para la comunidad científica. No vamos a detenernos en cuestiones técnicas, pero no está de más subrayar que, ya que se trata de facilitar las cosas a los colegas, el formato de los ficheros debe ser aquél o aquellos que resulten más fácilmente *importables* por los programas y aplicaciones más usados. De esta forma se simplifica la tarea de lectura e incorporación de la información.

Pero evidentemente, no basta con incluir un fichero de datos en una página Web. Para que sean útiles, tales datos han de estar adecuadamente documentados. Y con esto entendemos que ha de explicitarse qué variables son; en qué momento o momentos, cómo y con qué instrumentos la información ha sido recogida; qué transformaciones han sufrido los datos originales; cómo han sido seleccionados los casos; qué posibles niveles de agregación existen; etc. Toda esta información (que algunos denominan *metadatos*) debe añadirse a lo relativo a los resultados que ya hayan sido generados: cuáles se han obtenido y mediante el empleo de qué técnicas.

Por último, en lo que se refiere a las cuestiones de ética de la investigación, señalemos que, además del respeto a los acuerdos establecidos con los clientes o implicados, otro asunto muy importante es el que atañe a las garantías de confidencialidad, pues todo lo hasta aquí apuntado facilita los análisis secundarios, pero plantea graves problemas en lo que a esta cuestión se refiere.

No se trata de un conflicto insoluble, pero hay que prestar especial atención a los mecanismos y procedimientos tendentes a garantizar la confidencialidad bajo estos nuevos supuestos.

En primer lugar, la libre disponibilidad no está reñida con el establecimiento de filtros y niveles de acceso a la información que nos sirvan de razonable garantía de que nuestros datos van a ser usados con fines de investigación. Junto a este tipo de mecanismos propios de la seguridad informática, la propia disposición de los datos y ficheros puede también servir para garantizar la confidencialidad.

Aquí hay que prestar especial atención a aquellas variables que operan, o son susceptibles de operar, como identificadores. El problema no está tanto en los identificadores directos —pues lo normal es quitarlos— como en aquellas variables que tratadas

de forma conjunta pueden ser empleadas para identificar los casos. Según con qué muestras se opere y a qué niveles de agregación, con un cierto conocimiento de la población, esta posibilidad es bastante viable. Sieber (1998) describe, entre otros procedimientos, los que permiten introducir un error aleatorio en los datos de forma que no se interfiere con posibles análisis posteriores pero se dificulta o evita la identificación de casos o unidades.

Comunicación o presentación de resultados

Comenzábamos el apartado anterior enfatizando el carácter investigador de la evaluación de programas con objeto de justificar la difusión de los datos entre la comunidad científica. Ahora aquí procede poner el acento en lo que de *aplicada* tiene.

Y, como ya vimos, de ahí se deriva el hecho de que su objetivo primordial no sea tanto la búsqueda de la verdad o la producción de conocimiento, como el uso, el empleo de dicho conocimiento para la acción, para la mejora. A tal fin, los resultados han de ser presentados y comunicados a los clientes, implicados o interesados. En definitiva, a quienes tienen competencia, responsabilidad o capacidad decisoria de cara a la puesta en marcha de los procesos de intervención o de las acciones de mejora.

Si queremos que la evaluación produzca efectos, que sea realmente usada, es necesario que este proceso comunicativo sea eficaz. Y para ello es crucial tomar en consideración el hecho de que este tipo de audiencias es distinto de la comunidad científica a quien antes hemos hecho referencia.

Tal y como venimos haciendo a lo largo de este trabajo, no vamos aquí a extendernos sobre la comunicación del informe en general, sino que nos vamos a centrar en una parte del mismo: en la comunicación y presentación de *los resultados del análisis*. Y aquí la cuestión básica es conseguir que los mismos sean *inteligibles* para los destinatarios de la información.

Actualmente, el desarrollo metodológico pone a disposición del evaluador complejas técnicas analíticas para procesar los datos. Esta complejidad y versatilidad pueden permitir un mejor tratamiento de los datos, pero a su vez llevan aparejada una mayor dificultad de uso, y lo que más nos importa aquí, de comunicación de los resultados a audiencias no expertas.

Tal y como apunta Brinkerhoff (1983), no se trata, por descontado, de no emplear técnicas complejas con la excusa de que son difíciles de entender. Este autor advierte contra las sofisticaciones *innecesarias*, pero su primera recomendación es no hacer análisis simplistas habida cuenta de la complejidad de la realidad a evaluar.

En definitiva, el reto que tiene el evaluador no es nada fácil: hacer comprensibles a los destinatarios los resultados de los análisis que haya sido necesario emplear por complejos que sean, sin que esto suponga distorsión o pérdida significativa de información. Y éste es un tipo de comunicación que exige un esfuerzo y planteamiento muy distinto al de la difusión entre colegas en que se da por supuesto el conocimiento previo de la audiencia.

En cualquier investigación, después de obtener los resultados éstos han de ser *interpretados* tomando en consideración las características de las técnicas analíticas usa-

das y, sobre todo, confrontándolos con la realidad y el contexto en que se insertan. De esta interpretación, de esta búsqueda de *significado*, surgirán las respuestas a las hipótesis y cuestiones planteadas.

Pero en evaluación de programas, después de esta *primera vuelta* de análisis e interpretación, es necesario *re-analizar* dichos resultados para hacerlos comprensibles en la presentación. Y nótese que hemos empleado el término re-analizar con lo que queremos significar que este trabajo es parte también de la fase del análisis de los datos.

El énfasis que Patton (1997) pone en el uso de la evaluación, hace que las ideas que aporta en lo tocante a estas cuestiones nos resulten especialmente adecuadas y sugerentes. Como ilustración de lo dicho, en el capítulo dedicado al análisis y comunicación de resultados afirma que los resultados de un análisis de regresión pueden ser reducidos a tablas de contingencia y a un conjunto de estadísticos descriptivos de forma que en la presentación se focalice la atención en los resultados importantes.

Este alegato a favor de la simplicidad en la presentación (nada fácil de conseguir por otra parte) es completado con un ejemplo de la información presentada a los clientes respecto a una evaluación en la que, entre otras, se emplearon técnicas factoriales y de regresión para analizar los datos (pp. 317-321). Como conclusión afirma que los implicados no expertos pueden comprender e interpretar datos y resultados cuando son presentados en forma clara y legible usando gráficos y tablas estadísticas simplificadas.

Esto supone que la fase del análisis *no acaba* con la obtención e interpretación de los primeros resultados, de los resultados *brutos* que nos proporcionan los programas informáticos. Es competencia y trabajo (arduo) del investigador re-elaborar y re-analizar dichos resultados de forma que puedan garantizar una *comunicación eficaz*.

Y, además de Patton, son muchos los autores (con Tukey como ilustre precursor) que apuntan que la mejor manera de comunicar eficazmente es usando tablas y, sobre todo, *gráficos*. Pero para que la transmisión visual de información sea eficaz, no basta con apelar a la inteligencia espacial del observador confiando en que, por sí misma, será capaz de deducir de un gráfico los datos subyacentes y las relaciones que entre ellos se establecen. Si están mal diseñados pueden conseguir el efecto contrario: ocultar los datos (esto a veces se hace intencionadamente) o dispersar la atención del observador.

La capacidad de producir fácilmente gráficos cada vez más complejos se ha visto notablemente incrementada por las posibilidades que brindan la mayoría de los programas informáticos. Además de los programas orientados exclusivamente a este fin, la gran mayoría de los paquetes de análisis de datos incorporan potentes módulos gráficos con lo que la labor del investigador se ve notablemente facilitada.

De esta manera, es posible generar gráficos que cubren un amplio espectro de enfoques analíticos: desde la clásica descripción univariada o bivariada, hasta las tendencias, las series temporales o las nubes multivariables.

Pero al igual que el uso de un paquete estadístico no exime del conocimiento de la estadística, aquí también es preciso conocer las normas y criterios básicos que garantizan una comunicación visual eficaz. En nuestro campo, autores como Henry (1993,

1998) han estudiado el asunto y examinado los distintos tipos de gráficos y las condiciones que deben reunir.

Con respecto al contenido mismo de los gráficos la primera recomendación habitual estriba en que, además del gráfico en sí, en el mismo deben aparecer datos. El objetivo es claro: el gráfico actúa como soporte y la información textual o numérica permite *leer* e interpretar correctamente la información iconográfica. En conclusión, los gráficos deben ser ricos en información mostrando la complejidad subyacente a los datos. Pero a la vez, conviene no recargar excesivamente cada gráfico pues dificulta su lectura. En vez de intentar mostrar toda la información en un solo gráfico, es mejor emplear varios: por ejemplo, una serie para las descripciones univariadas y otra para mostrar las relaciones que pudieran existir entre variables. En el delicado y difícil equilibrio entre estos dos polos radica una de las claves de una comunicación visual eficaz.

Pero además de las cuestiones de contenido, al hablar de información gráfica no podemos obviar lo tocante a las propias formas, a las cuestiones de formato. En palabras del propio Henry (1998: p. 537), aquí hay cuestiones tan importantes que no conviene dejarlas tal y como aparecen en las opciones *por defecto* pre-establecidas por los programas. Entre ellas cita lo relativo a la orientación, los ejes, las tramas y el relleno o las proporciones. Con el desarrollo actualmente alcanzado por monitores e impresoras, nosotros nos atrevemos a añadir el tratamiento del color. El conocimiento de las normas básicas derivadas de la teoría del color, la complementariedad y la saturación, creemos que contribuirían al menos a evitar gráficos que —por la gama de colores usados— no sólo son de gusto estético más que discutible sino que —lo que es más grave— dificultan cuando no impiden la aprehensión de los datos.

El resto de las posibilidades que brinda la informática gráfica (animación, efectos tridimensionales, etc.) están contribuyendo al auge y desarrollo del campo de la *visualización de los datos*. Por razones obvias no podemos extendernos más en estas cuestiones. Además del ya citado Tukey, la obra de Bertin (1973), *Sémiologie graphique* es todo un clásico en la materia. Junto a ellas, nos permitimos recomendar a los interesados las obras de Wallgren y otros (1996) y de Jacoby (1998).

Muy útil y completa es la página Web sobre visualización de datos del *Statistical Consulting Service* de la universidad canadiense de York. Su dirección es la siguiente: *http://www.math.yorku.ca/SCS/Gallery/*. En la misma aparecen ejemplos de buen y mal uso de los gráficos, además de muchos enlaces con páginas dedicadas a los gráficos y a la visualización de datos.

BIBLIOGRAFÍA

Alvira, F. (1991): *Metodología de la evaluación de programas*. Madrid: CIS.

Apodaca, P. M. (1999): Evaluación de los resultados y del impacto. *Revista de Investigación Educativa*, 17, 2, 363-377.

Arnau, J.; Ruiz, M. y Salvador, F. (1988): Evaluación del tratamiento en diseños conductuales mediante técnicas de análisis de series temporales: una aproximación computacional. *Evaluación Psicológica*, 4, 1, 51-62.

370

Batista J.M. y Coenders G. (2000): *Modelos de ecuaciones estructurales*. Madrid: La Muralla, Hespérides.

Bertin, J. (1973): *Sémiologie graphique*. Paris: Mouton.

Bickman, L. y Rog, D. J. (1998): Introduction: Why a Handbook of Applied Social Research Methods? en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, ix-xix.

Botella, J. y Barriopedro, Mª. I. (1995): Análisis de datos. En Fernández-Ballesteros, R. (Ed.). *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis, 173-203.

Brinkerhoff, R.D. y otros (1983): *Program Evaluation. A practicioner´s guide for trainers and educators*. La Haya: Kluwer-Nijhoff.

Bryk, A., Raudenbush, S.W. (1992): *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Carballo, R. (1990): Algunas aplicaciones del análisis de correspondencias a la interpretación de tablas de contingencia en la evaluación de un programa de formación profesional ocupacional. *Revista de Investigación Educativa*, 8, 16, 537-550.

Castro, M. (1999): Modelos multinivel aplicados al meta-análisis. *Revista de Investigación Educativa*, 17, 2, 445-453.

Castro, M. y De la Orden, A. (1998): Análisis y comparación de las características de los principales modelos de síntesis meta-analítica: aportaciones de los modelos jerárquicos lineales. *Bordón*, 50, 2, 119-134.

Cohen, J. (1988): *Statistical power analysis in the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H.M. (1981): On the effects of significance and the significance of effects. *Journal of Personality and Social Psycology*, 41, 1013-1018.

Cornejo, J.M. y Martínez, J.J. (1992): Análisis de datos e intervención social en la infancia. *Anuario de Psicología*, 53, 47-59.

De la Orden, A. (1985): Investigación evaluativa. En De la Orden, A: *Investigación Educativa. Diccionario de Ciencias de la Educación*. Madrid: Anaya, 133-137.

De la Orden, A. (1999): Los modelos jerárquicos lineales y su aplicación a la investigación educativa. *Revista de Investigación Educativa*, 17, 2, 425-427.

De la Orden, A. y Martínez de Toda, M. (1992): Metaevaluación educativa. *Bordón*, 43, 4, 517-527.

De Miguel, M. (1999): La evaluación de programas: entre el conocimiento y el compromiso. *Revista de Investigación Educativa*, 17, 2, 345-348.

Echeburua, E. y otros (1996): Tratamiento cognitivo-conductual del trastorno de estrés postraumático crónico en víctimas de maltrato doméstico: un estudio piloto. *Análisis y Modificación de Conducta*. 22, 85, 627-654.

Escofier B. y Pagès J. (1990): *Analyses factorielles simples et multiples*. Paris: Dunod.

Fernández-Ballesteros, R. (Ed.). (1995): *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.

Fetterman, D.M. (1998): Ethnography, en Bickman, L. y Rog, D. J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 473-504.

Filella, G. (2000): Evaluación del programa de orientación para la inserción laboral en los centros penitenciarios. *Revista de Investigación Educativa*, 18, 1, 83-96.

Francis, E. y otros (1999): Improving cost-effectiveness in a substance abuse treatment program. *Psychiatric Services*, 50, 5, 633-635.

Freixa, M. y otros (1996): Una propuesta de entrevista semi-estructurada para la evaluación de futuros padres adoptivos: aplicación y estudio mediante el análisis textual. *Anuario de Psicología*, 71, 37-49.

Gaviria, J.L. (1999): El enfoque multinivel en la evaluación de sistemas educativos. *Revista de Investigación Educativa*, 17, 2, 429-443.

Glass G.V., Mcgaw B. y Smith M.L. (1981): *Meta-analysis in social research*. Thousand Oaks, CA: Sage.

Goldstein, H. (1995): *Multilevel Statistical Models*. London: Edward Arnold; New York: Halsted Press.

González Ramírez, T. (2000): Metodología para la enseñanza de las matemáticas a través de la resolución de problemas: un estudio evaluativo. *Revista de Investigación Educativa*, 18, 1, 175-199.

Hannan, P.J. y Murray, D.M. (1996): Gauss or Bernoulli? A Monte Carlo Comparison of the performance of the linear mixed-model and the logistic mixed-model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Review*, 20, 3, 338-352.

Hair J.F., Anderson R.E., Tatham R.L. y Black W.C. (1999): *Análisis multivariante*. Madrid: Prentice Hall Iberia.

Hedges L.V. y Olkin I. (1985): *Statistical methods for meta-analysis*. London: Academic Press.

Henry, G.T. (1993): Using graphical displays for evaluation data. *Evaluation Review*. 17, 60-78.

Henry, G.T. (1998): Graphing Data, en Bickman, L. y Rog, D. J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 527-556.

Hernández, P.; Aciego, R. y Domínguez, R. (1994): Crecimiento personal y educación en valores sociales: evaluación del programa PIECAP. *Revista de Psicología General y Aplicada*, 47, 3, 339-347.

Herraiz Gonzalo, F. (1998): Revisión de un programa cognitivo-conductual en el departamento de Psiquiatría del Centro Penitenciario de Hombres de Barcelona. *Psiquis*, 19,10, 425-432.

Hunter J.E., Schmidt F.L. (1990): *Methods of meta-analysis*. Newbury Park: Sage.

JACOBY, W.G. (1998): *Statistical Graphics for Visualizing Multivariate Data*. Thousand Oaks, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1994): *The Program Evaluation Standards*. Thousand Oaks, CA: Sage.

Knight, C.C. y Kuleck, W.J. (1999): The value of multimethod qualitative/quantitative research methodology in an educational program evaluation: a case study. *Mid-Western Educational Researcher*, 12, 2, 6-14.

Kreft, I.G.G. (1998): An illustration of item homogeneity scaling and multilevel analysis techniques in the evaluation of drug prevention programs. *Evaluation Review*, 22, 1, 46-77.

372

Krull, J.L. y Mackinnon, D.P. (1999): Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 4, 418-444.

LAVIT C. (1988): *Analyse Conjointe de Tableaux Quantitatifs*. Paris: Masson.

Lazaro Martínez, A. (1992): La formalización de indicadores de evaluación. *Bordón*, 43, 4, 477-494.

Lebart L., Morineau A. y Piron M. (1995): *Statistique exploratoire multidimensionnelle*. Paris: Dunod.

Linn R.L. (2000): Assesments and Accountability. Disponible en: http://www.aera.net/pubs/er/arts/29-02/linn01.htm

Lipsey M.W. (1990): *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Mcneal, R.B. y Hansen, W.B. (1995): An examination of strategies for gaining convergent validity in natural experiments. *Evaluation Review*, 19, 2, 141-158.

Murillo, F.J. (1999): Los modelos jerárquicos lineales aplicados a la investigación sobre eficacia escolar. *Revista de Investigación Educativa*, 17, 2, 453-458.

Patton, M.Q. (1997): *Utilization-Focused Evaluation. The New Century Text (3rd ed.)*. Thousand Oaks, CA: Sage.

Pérez, M. y Salvador, L. (1998): Evaluación experimental del proyecto de inteligencia Harvard. *Papeles del Psicólogo*, 71, 33-39.

Pérez Juste, R. (1995): Metodología para la evaluación de programas educativos. En Medina, A. y Villar, L.M. (coord.) *Evaluación de programas educativos, centros y profesores*. Madrid: Universitas, 109-144.

Plotnick, R.D. y Deppman, L. (1999): Using benefit-cost analysis to assess child abuse prevention and intervention programs. *Child Welfare*, LXXVIII, 3, 381-407.

REPETTO, E. y otros (1994): Últimas aportaciones en la evaluación del programa de orientación metacognitiva de la comprensión lectora. *Revista de Investigación Educativa*, 23, 314-323.

Sáenz O. y González A. (1990): Aplicación de cadenas Markovianas a los procesos de intervención educativa. *Revista de Investigación Educativa*, 16, 2, 493-499.

Sánchez Algarra, P. y Anguera Argilaga, M.T. (1993): Aproximación al PERT en evaluación de programas desde las técnicas matemáticas de análisis de grafos. *Anales de Psicología*, 9, 2, 213-226.

Sieber, J.E. (1998): Planning Ethically Responsible Research, en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 127-156.

Sogunro, O.A. (1997): Impact of training on leadership development. *Evaluation Review*, 21, 6, 713-737.

Tejedor, F.J.; García-Valcárcel, A. y Rodríguez Conde, Mª. (1994): Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo. *Revista de Investigación Educativa*, 23, 93-127.

Tukey, J.W. (1977): *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Uriel, E. (1992): *Análisis de series temporales. Modelos ARIMA*. Madrid: Paraninfo.

Wallgren, A. y otros (1996): *Graphing Statistics and Data*. Thousand Oaks, CA: Sage.

Wang, J. y otros (1999): Evaluation of HIV risk reduction intervention programs via latent growth model. *Evaluation Review*, 23, 6, 648-662.

Ximénez C. y San Martín R. (2000): *Análisis de varianza con medidas repetidas*. Madrid: La Muralla, Hespérides.

Yin, R.K. (1998): The Abridged Version of Case Study Research: Design and Method, en Bickman, L. y Rog, D.J. (Eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 229-259.

DIRECCIONES INTERNET:
http://www.ioe.ac.uk/multilevel
http://www.math.yorku.ca/SCS/Gallery/
http://www.aera.net/pubs/er/arts/29-02/linn01.htm

371

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 381-389

LA FUNCIÓN OPTIMIZANTE DE LA EVALUACIÓN DE PROGRAMAS EVALUATIVOS*

Arturo de la Orden Hoz
Universidad Complutense de Madrid

Es evidente que en el mundo actual y, especialmente, en las sociedades más desarrolladas, la evaluación ocupa un lugar tan amplio y destacado que podríamos definir esta etapa histórica como «la era de la evaluación». Para explicar esta situación satisfactoriamente, habríamos de recurrir, sin duda, a diferentes vías y considerar muy diversos factores. Sin embargo, una constatación de las condiciones generales en que este hecho se produce nos permitirá un primer acercamiento a su explicación. En efecto, la evaluación es un fenómeno que se desarrolla y adquiere relevancia en momentos históricos caracterizados por el cambio social. En realidad, se trata de un mecanismo permanente de ajuste de las decisiones en las actividades sociales organizadas, cuya función es optimizar las alternativas. Si no hubiera cambio, no habría alternativas y, en consecuencia, la evaluación perdería su significado.

La evaluación como característica inherente a toda actividad humana intencional, se manifiesta tanto en la vida cotidiana, escasamente planificada, cuanto en la acción sistemática de las empresas individuales o colectivas del hombre. En el devenir de la rutina diaria, el hombre evalúa, valora constantemente personas, objetos, situaciones y acontecimientos, juzgándolos, previa contrastación con instancias o modelos, más o menos explícitamente representados.

En la acción individual o colectiva sistematizada (la política, la administración, la actividad económica, la medicina, etc.) la evaluación aparece como un componente esencial. La valoración de las consecuencias, el control de la calidad, el diagnóstico de la salud, que se utilizan como indicadores del proceso o de la adecuación de los resultados, son diferentes manifestaciones del fenómeno evaluativo.

---

* Este artículo está apoyado en otros trabajos del autor (De la Orden, 1982 y 1997).

Esta multiplicidad de situaciones evaluativas justifica la visión de este campo de actividad como un conjunto disciplinar relativamente disperso, cultivado como ocupación parcial por profesionales de distintas disciplinas para satisfacer necesidades, a su vez, dispersas, pero que, como afirma Morell (1990), comportan un núcleo de principios, doctrinas o creencias precisas, pero mal definidas. En la evaluación podemos rastrear elementos procedentes de las disciplinas de origen de los evaluadores —pedagogía, psicología, sociología, economía, política, etc.— cuya formación básica está vinculada a su área de procedencia, pero, en general, sienten una necesidad común de definir su afiliación profesional, al menos, parcialmente, en términos no estrictamente específicos del campo disciplinario en que se mueven.

No obstante, en los últimos 25 años, la evaluación ha sufrido un proceso de transformación profunda de sus bases estructurales y conceptuales. Como señala House (1992), estructuralmente se ha ido integrando cada vez más en el funcionamiento organizativo de las instituciones y contextualmente se ha flexibilizado, aceptando un alto grado de pluralismo teórico y metodológico.

Hoy, en un ambiente de cambio educativo generalizado, efecto tanto de las grandes reformas en marcha como de la innovación permanente, a nivel institucional y de aula, la evaluación educativa cobra también su máximo protagonismo y presenta características similares a las apuntadas para el campo en su conjunto, como demuestra el énfasis puesto en la evaluación del aprendizaje de los alumnos, la evaluación de los sistemas educativos en la perspectiva regional, nacional e internacional, los avances en la evaluación de centros escolares y profesores y especialmente la evaluación de programas educativos específicos y de formación continua de profesionales.

HACIA UNA PRECISIÓN CONCEPTUAL

Como premisa para la caracterización de la función optimizante de la evaluación de programas educativos, objetivo de este artículo, parece pertinente delimitar con la máxima precisión el propio concepto de evaluación como referente de los argumentos y análisis conducentes a la demostración de la tesis planteada.

La evaluación constituye esencialmente un juicio de valor sobre una realidad y, como todo juicio, se apoya en una comparación. Comparación y juicio son, pues, los componentes esenciales de la evaluación, su núcleo conceptual. Como la comparación exige dos términos, para evaluar precisamos la determinación de la instancia o modelo con la cual se ha de comparar la realidad objeto de evaluación, es decir, el patrón o criterio. La comparación implica, en consecuencia, de una parte el conocimiento, lo más completo posible, de la realidad (programas en nuestro caso) a evaluar, que necesariamente se apoya en un proceso previo de información (recogida de datos, medidas, etc.), y de otra, la descripción ajustada del término de comparación, del patrón, del criterio.

Por otra parte, la evaluación tiene un carácter instrumental. En educación, como en otros campos, evaluamos para algo que trasciende a la propia evaluación. Se evalúa para promocionar a los alumnos, o a los profesores, para adoptar tal o cual objetivo, programa o procedimiento didáctico, para mejorar, en suma la enseñanza y el aprendi-

zaje. La evaluación, pues, cumple unas funciones definidas en el contexto educacional; sus resultados son susceptibles de un uso determinado, de acuerdo con las funciones previstas. En general, podemos afirmar que se evalúa para tomar decisiones respecto al proceso o producto evaluados. Al juicio de valor en el que la evaluación consiste sigue una decisión.

En consecuencia, podríamos definir la evaluación de programas como el proceso sistemático de recogida, análisis e interpretación de información relevante y fiable para describir un programa educativo, o una faceta significativa del mismo, y formular un juicio sobre su adecuación a un criterio o patrón, que represente un valor aceptado, como base para la toma de decisiones sobre tal programa o faceta programática.

LAS FUNCIONES DE LA EVALUACIÓN

Dado el carácter instrumental de la evaluación, ya señalado, la determinación de las funciones y usos constituye una de las exigencias básicas, tanto en la perspectiva teórica —su conceptualización— como en la perspectiva de la práctica evaluativa. Cual sea la concepción del sentido último, del para qué de la evaluación y cual sea su concreción en los objetivos específicos de cada proyecto evaluativo, son factores altamente condicionantes, quizá determinantes, del efecto de la evaluación sobre el objeto evaluado y su contexto, es decir, del éxito o fracaso del instrumento evaluativo.

Convencionalmente se habla, por una parte, de la función de diagnóstico como base de la intervención educativa y de pronóstico y prevención; y, por otra, de la función de control y rendimiento de cuentas que supone informar a usuarios y consumidores de programas educativos, a las familias, a los gestores y decisores y, en último extremo, a la sociedad en general, acerca del valor o mérito del objeto evaluado —programas educativos— y certificar conocimientos y competencias generales y profesionales adquiridas en tales programas por los participantes, con garantía de objetividad.

Tradicionalmente estas dos categorías de usos de la evaluación han sido denominadas función formativa y función sumativa respectivamente. El primero en usar esta terminología fue Scriven (1967) quien caracterizó a la evaluación formativa como evaluación de procesos y a la evaluación sumativa como evaluación final, asumiendo, claro está, que la evaluación final de un programa se identifica con la evaluación de su producto. Poco después, Bloom (1969 y 1971) circunscribió la función formativa a la evaluación diagnóstica que, refiriéndose a los alumnos, trataría de determinar el estado de su aprendizaje en cualquier momento del proceso discente, mientras que la función sumativa sería la propia de una evaluación centrada en situar a cada alumno, respecto a su aprendizaje, en el punto que le corresponde en un continuo de competencia (del nivel más bajo —incompetencia— al nivel más alto —máxima competencia—). Finalmente, Stuffelebeam (1974) matiza estos conceptos asignando a la evaluación formativa el carácter de pro-activa y cuya función podría identificarse con la toma permanente de decisiones sobre los procesos del programa; y considerando a la evaluación sumativa como retro-activa con la función de exigir responsabilidades o rendición de cuentas, naturalmente como consecuencia de los resultados finales del programa.

Las diferencias entre ambos tipos de función responden básicamente a una doble consideración: el momento de su aplicación en el continuo del proceso educativo, por un lado, y la modalidad de las decisiones a que da lugar, es decir, el uso que se hace de la evaluación en cada caso, por otro lado. Ambas perspectivas combinadas determinan, a su vez, la faceta del programa en que con más probabilidad se centrará la evaluación (contexto, plan, proceso o producto).

En consecuencia, la evaluación de carácter formativo se centra en las fases previas al desarrollo del programa (la determinación de las necesidades a satisfacer, en el desarrollo mismo, en su plan y diseño y en los distintos estadios de su puesta en operación, lo que supone decisiones diagnósticas y de retroinformación a los planificadores y aplicadores del programa para el perfeccionamiento del mismo.

Más precisamente, la evaluación formativa proporciona la información adecuada para la mejora de los procesos educativos y a través de ellos, del producto.

Por su parte, la evaluación sumativa tiene generalmente carácter final y por tanto se centra básicamente en los resultados inmediatos, productos mediatos e impacto de los programas educativos.

Las decisiones que definen esta función evaluativa apuntan a la selección, clasificación y promoción de alumnos y a la certificación de conocimientos y competencias. Así mismo constituye la forma típica de evaluación para decidir la continuidad o terminación de un programa, modelo de organización, centro, etc. En resumen, controlar el valor de un programa (funcionalidad o pertinencia, eficacia y/o eficiencia).

Recientemente, y casi siempre en conexión con las discusiones en torno a los modelos de evaluación de instituciones educativas, se suele aceptar, en general de forma mecánica y acrítica, la existencia de dos modalidades de evaluación, la orientada al control, identificada en gran medida con la evaluación sumativa, y la orientada a la mejora, identificada con la evaluación formativa. La frecuente referencia al equiparar la evaluación formativa a la mejora y perfeccionamiento del proceso, y las escasas alusiones a decisiones de este tipo al caracterizar a la evaluación sumativa, puede ser una explicación plausible. No parece necesario documentar, por obvio, que la modalidad evaluativa orientada a la mejora es la que goza de una aceptación casi universal.

Sin embargo, la objeción más suave a esta posición es la de su simplicidad. La ausencia de matices a su vez convierte esta clasificación en un error. Una consideración atenta del concepto de evaluación y sus implicaciones pone claramente de manifiesto que la evaluación, toda evaluación, se justifica en su condición instrumental por su función optimizante del objeto evaluado y su contexto. Si esto es así, la evaluación orientada al control solo se justificará como evaluación educativa si mejora la educación. En esta perspectiva, la decisión de suprimir un programa, una disciplina, un plan de estudios e incluso una institución escolar como consecuencia de una evaluación sumativa o de control, se supone que mejora la educación. La diferencia con la evaluación formativa (de mejora) es que el perfeccionamiento, el proceso de optimización, es diferente, está pautado por decisiones distintas.

En otro trabajo (De la Orden, 1981) formulaba así lo que podría ser la base conceptual de la función general de la evaluación educativa propuesta, es decir, la optimización de la educación: Sea cual fuere la modalidad o tipo de evaluación adoptada en

378

cualquier contexto educativo, su influencia sobre el proceso y el producto de la educación es decisiva y afecta directa o indirectamente a su calidad, siempre claro está que la evaluación tenga consecuencias, es decir, siempre que se tomen las decisiones en función de las cuales se decidió la evaluación. Una evaluación sin consecuencias es inútil y puede ser perjudicial para el objeto evaluado y su contexto.

INFLUENCIA DE LA EVALUACIÓN EN LA CALIDAD DE LA EDUCACIÓN

La explicación de la influencia constatada en la evaluación en el proceso y el producto de la educación se apoya en el hecho de que las necesarias consecuencias de la evaluación modifican en algún sentido el objeto evaluado, en nuestro caso, el programa educativo, y también en alguna forma, la situación de las audiencias implicadas en el mismo. Estos efectos al ser percibidos por las audiencias como positivos o negativos —deseables o no deseables— condicionan el alto grado la conducta de estas audiencias en el contexto del programa educativo. Por otra parte, el elemento determinante del sentido de la modificación del objeto evaluado es, sin duda, la instancia de referencia. En efecto, los criterios indican lo que se exige del programa, del profesor, del alumno, para que el juicio de valor sobre el mismo sea favorable. La percepción por las audiencias críticas de esta situación guiará su conducta en la medida en que están interesadas realmente en el objeto evaluado y en las consecuencias de su evaluación. Parece evidente que los alumnos intentan satisfacer las exigencias (criterios) de los exámenes respecto a su aprendizaje; los profesores, en la medida en que la evaluación tenga consecuencias para su estatus y su profesión, tratarán de ajustar su docencia a las exigencias —criterios— del modelo evaluativo; los gestores de cualquier programa educativo tratarán de modelar los procesos y productos del mismo a las exigencias —criterios— de la evaluación. De forma general, por tanto, podemos afirmar que el modelo de comportamiento implícito en los criterios de evaluación, al constituir una expresión concreta de lo que se espera del programa realmente, puesto que es lo que exige de hecho, define de una manera efectiva los objetivos reales y operantes de la educación independientemente de lo estipulado formalmente y de que existan o no formulaciones expresas de los mismos.

En consecuencia, si consideramos el principio de la determinación teleológica de los medios (con toda la relatividad que se quiera, los fines y objetivos condicionan de alguna manera los medios para alcanzarlos), debemos concluir que la evaluación, al prescribir realmente los objetivos de la educación, determina, en gran medida, las características de la intervención y el aprendizaje resultante, lo que los alumnos aprenden y cómo lo aprenden, lo que los profesores enseñan y cómo lo enseñan, los contenidos y los métodos, en otras palabras, el producto y el proceso del programa. En efecto, si la calidad de la educación está vinculada a un sistema de coherencias en el proceso total de la educación, la virtualidad de la evaluación como estímulo o freno de la calidad educativa dependerá enteramente de las características del esquema evaluador: Si los criterios y los modos de evaluación, que determinan a su vez los objetivos reales, y en última instancia el producto del programa, son coherentes con los objetivos formalmente establecidos —o sobre los que existe un consenso implícito—, y, a través de

ellos, con el sistema de valores del que derivan, la evaluación actuará como el más poderoso factor de promoción de la calidad educativa al garantizar la congruencia y eficacia del sistema en su conjunto. Si, por el contrario, los criterios y modos de evaluación movilizan los procesos educativos hacia los objetivos no coherentes con las metas formalmente establecidas o implícitamente aceptadas del programa ni, en consecuencia, con el sistema axiológico en que se sustentan, la evaluación se constituye en el mayor obstáculo a la calidad.

CONDICIONES GENERALES DE LA EVALUACIÓN COMO INSTRUMENTO DE OPTIMIZACIÓN EDUCATIVA

Dado que la evaluación del aprendizaje puede influir positiva o negativamente en la calidad de la educación, resulta necesario determinar qué características debe poseer un modelo evaluativo para maximizar sus efectos positivos y minimizar o anular los negativos.

Evidentemente, la condición fundamental de un sistema de evaluación es que sea educativamente válido, lo que implica no solamente que los instrumentos de medida utilizados (test, pruebas, escalas de observación, procedimientos no formales, etc.) sean válidos en el sentido tradicional, sino también que exista un máximo de congruencia entre los criterios y modos de evaluación, la especificación de los objetivos del programa y el proceso de intervención. El elemento común entre criterios de evaluación, objetivos y proceso de intervención es el comportamiento cognitivo, afectivo y psicomotor de los alumnos claramente especificado. En efecto, los objetivos no son otra cosa que la especificación de este comportamiento o de la capacidad para el mismo en forma de resultados esperados del proceso educativo; los criterios de evaluación, lo que exige, constituyen, en conjunto, una muestra representativa de tal comportamiento; y el proceso didáctico se apoya en este comportamiento como base para su adquisición.

En consecuencia, la validez educativa de la evaluación del aprendizaje, como base de la calidad de la educación, significa:

1) Que los objetivos del programa sean claramente identificados y formulados sin ambigüedad. Para que la evaluación no se desvíe de los objetivos del programa, es necesario conocer con la máxima precisión tales objetivos.
2) Que los criterios de evaluación (lo que se exige del programa y cómo se exige) deben constituir una adecuada muestra representativa de los contenidos y conductas especificadas en los objetivos. Esta es la condición básica de la evaluación como palanca de la calidad de la educación. Todo profesor ha comprobado que los estudiantes aprenden mucho más y de manera más efectiva todo aquello que es objeto de evaluación y control. La evidencia experimental ha confirmado claramente esta hipótesis. Baste citar el ya antiguo estudio comparativo de Walker y Schaffarzick (1974) sobre evaluaciones del currículum en el que se puso de manifiesto que cuando las pruebas de rendimiento están en línea con las metas del programa innovador éste resulta más eficaz que el programa tra-

dicional con el que se compara. Por el contrario, cuando las pruebas no cumplen esta condición, resultan superiores los programas tradicionales. Por otra parte, el intento de alinear los criterios de evaluación con los objetivos contribuye decisivamente a clarificar a éstos y a resaltar su importancia en el pensamiento de profesores y alumnos. Las metas educacionales se hacen más definidas y significativas, incrementando la coherencia entre criterios de evaluación, objetivos formales y aspectos del aprendizaje considerados por alumnos y profesores como resultados educativos valiosos e importantes.

3) Que la forma, procedimientos e instrumentos de evaluación exijan los comportamientos especificados en los objetivos del modo más directo. Es decir, la técnica evaluativa empleada (observación de procesos, análisis de productos, informe, examen escrito tradicional, cuestiones de recuerdo o reconocimiento, pruebas con materiales o sin materiales, exámenes orales, etc.) debe estar directa y estrechamente relacionada con las características del aprendizaje o realización deseadas. Si un objetivo de un programa de Química es que los estudiantes manejen adecuadamente determinadas piezas del equipo del laboratorio, exigir la descripción oral de cómo manejaría tales piezas, o la elaboración de un esquema gráfico del proceso, son formas evaluativas menos válidas que requerir el desarrollo real del proceso en el propio laboratorio.
4) Que la evaluación sea fiable y objetiva en el sentido de que el azar o los errores instrumentales tengan un efecto mínimo en los resultados.

Una evaluación así concebida supone, en principio, el requisito esencial para asegurar el sistema de coherencias en que consiste la calidad de la educación. La relación directa y la correspondencia entre, objetivos formales, objetivos reales, procesos instructivos y resultados de un programa, de tal manera que resulten isomórficos, constituye la única garantía de un sistema con un alto grado de coherencia interna y externa con las expectativas y los resultados.

OPTIMIZACIÓN EDUCATIVA Y EVALUACIÓN DE LAS FASES DEL PROGRAMA

Una forma de incrementar la potencia optimizante de los programas educativos consiste en asociar la evaluación a su planificación, desarrollo y aplicación. Una de las características básicas de la evaluación es su carácter instrumental. La evaluación, como señalamos, más allá de un juicio de valor, se justifica porque permite el ajuste de las decisiones sobre el objeto o faceta del objeto evaluado, y este ajuste decisional supone en todas las circunstancias una mejora del objeto (programa, currículum, innovación de todo orden). La función evaluativa, como hemos señalado, cualquiera que sea el objetivo específico en cada caso, será siempre la optimización de las alternativas a la situación evaluada.

Cualquiera que sea el origen, objetivos, estructura y alcance del programa educativo, el paso inicial en evaluación consiste en determinar su necesidad. El análisis del contexto llevará a la selección de problemas, al diagnóstico situacional, que se plasmará en un conjunto de necesidades educativas priorizadas que habrán de ser satisfechas

381

a través del programa. La evaluación de necesidades se convierte en la fase previa de la planificación ya que su resultado será la identificación de las metas de la acción educativa. De esta forma, la evaluación de necesidades asegura la funcionalidad del programa –dimensión básica de la calidad— al facilitar su coherencia con las aspiraciones, expectativas y necesidades de individuos y grupos.

El proceso de planificación y desarrollo del programa debe ser también objeto de evaluación. En este sentido habrá de ser valorado el modelo de programa adoptado y el propio modelo de planificación y desarrollo que se sigue, así como la política de estímulo y de planificación de programas innovadores.

El producto de la planificación y desarrollo del programa es el plan o propuesta de acción que, a su vez, habrá de ser sometido a evaluación, antes de su puesta en operación. Esta fase de la evaluación debe asegurar la adecuación del diseño del programa y su coherencia interna, es decir, la congruencia inter e intra componentes del plan: objetivos, contenidos, estrategias, experiencias y actividades de aprendizaje y ambientes de intervención, materiales de apoyo y previsiones de evaluación. Un plan coherente es una condición necesaria, aunque no suficiente, de una intervención coherente, exigencia fundamental de la calidad educativa.

La evaluación de la puesta en operación del plan constituye una fase esencial para asegurar la calidad del programa y por tanto, la optimización educativa. Se trata de la evaluación de los procesos educativos generados por la aplicación del programa, a través de la observación, de pruebas piloto y seguimiento. La evaluación del proceso permitirá constatar en que grado es innovadora la conducta docente y discente y permitirá tomar decisiones adecuadas en orden a evitar que la inercia desdibuje el perfil de la innovación, como ha sido comprobado en diversos estudios. (Alexander, Murphy y Woods, 1996; Ekholm y Kull, 1996) Así mismo, la evaluación de los procesos educativos reales durante el periodo de ejecución del plan constituye un requisito imprescindible para la evaluación del producto del programa. Es absolutamente necesario saber si los resultados constatados se deben a los procesos planificados o a las formas de conducta, que sólo aparentemente se presentan como las exigidas por el programa.

La evaluación del producto o resultados educativos del programa apuntan directamente a determinar, por un lado, la eficacia, y por otro, la eficiencia del mismo y, por tanto, la optimización educativa. La eficacia del programa se constata bien en términos absolutos —grado en que el producto y el impacto de la acción educativa coinciden con los objetivos y productos previstos—, bien en términos relativos —grado en que los resultados de los procesos programados se acercan más o menos a los productos o resultados de los programas convencionales de educación o de otros programas diferentes—.

La evaluación, en suma, permite detectar los problemas educativos, su extensión, complejidad y relevancia a medida que se van produciendo. Esta constatación puede dar lugar a la planificación de cambios e innovaciones en programas concebidos como alternativas de solución a tales problemas. Una vez puesto en marcha el plan de trabajo, como hipótesis de solución, la evaluación formativa proporciona, desde el primer momento, información que permite juzgar en qué grado el programa o el cambio pla-

382

nificado produce los efectos previstos en el aprendizaje de los alumnos; y sugiere las eventuales modificaciones necesarias para el logro de los objetivos previstos.

En resumen, la evaluación de programas educativos, en la medida en que los criterios sean congruentes con los objetivos previstos y el diseño de evaluación sea el adecuado constituye la palanca más importante para canalizar los procesos y los productos educativos hacia objetivos valiosos y deseables social e individualmente y, por tanto, promueve la funcionalidad, la eficacia y la eficiencia de la educación y, en consecuencia, su calidad.

BIBLIOGRAFÍA

Alexander, P.A., Murphy, P.K. y Woods, B.S. (1996). *Of the squalls and fathoms: Navegating the seas of educational innovation*. Educational Rearcher. 25 (3) 31-39.

Bloom, B.S. (1969). *Some theoretical issues relating to educational evaluation*. En R.W. TYLER (Ed.) «Educational Evaluation: new roles, new means». 68th Yearbook of the N.S.S.E. Chicago. University of Chicago Press.

Bloom, B.S. y otros (1971). *Handbook on formative and sumative of student learning*. N. York, McGrawittill.

De la Orden, A. (1982). *La Educación Educativa*. Buenos Aires, Docencia.

De la Orden, A. (1997). *Evaluación y optimización educativa*. En H. Salmeron y otros (Eds.). Evaluación Educativa. Granada, Grupo Editorial Universitario.

De la Orden, A. (1981). *Evaluación del aprendizaje y calidad de la educación*. En varios «La calidad de la educación». Madrid, C.S.I.C. Instituto de Pedagogía.

Ekholm, M. y Kull, M. (1996). *School Climate and Educational Change*. EERA Bulletin 2 (2) 3-11.

House, E.R. (1992). *Tendencias en evaluación. Revista de Educación*, nº 299, p. 43.

Scriven, M. (1967). The methodology of evaluation. AERA Monograph Series on Currículum Evaluation, nº1, pp. 39-83.

Stufflebeam, D.L. (1974). *Metaevaluation*, The Evaluation Center, Kalamazoo, Western Michigan University Press.

Walter, D.F. y Schaffarzick, J. (1974). *Comparing Curricula*. Review of Educational Research, nº 44, pp. 83-111.

SEGUNDA PARTE

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 393-404

EL MUESTREO Y SU PROBLEMÁTICA EN LAS EVALUACIONES DE PROGRAMAS INSTITUCIONALES

José Luis Gaviria Soto
Universidad Complutense de Madrid

1. INTRODUCCIÓN

Cuando se aborda la evaluación de resultados en los programas institucionales, especialmente en los niveles básicos de la educación, se siente siempre la tentación de evaluar a todos y cada uno de los individuos sobre los que actúa el sistema. Esta posibilidad puede estar en ocasiones justificada. Se trata del caso en el que la población de referencia es tan pequeña que la satisfacción de ciertos niveles de precisión impone una muestra cuyo tamaño es no mucho más pequeño.

Sin embargo por regla general la extracción de muestras presenta numerosas ventajas que hacen recomendable y beneficioso el esfuerzo. Siguiendo a Som (1996), podemos afirmar que las técnicas de muestreo tienen la ventajas de:

— mayor economía,
— intervalos de tiempo más cortos entre las distintas recogidas de información,
— más amplio espectro de la información,
— mayor calidad del trabajo,
— y posibilidad de la valoración de la fiabilidad de los resultados

Ciertamente si nos centramos en un grupo comparativamente pequeño respecto de la población, podremos recoger más información de cada individuo, asegurarnos que los datos obtenidos de cada sujeto son fiables, llevar a cabo la recogida de información más frecuentemente, y todo ello a un menor coste.

Pero por supuesto los procedimientos de muestreo pueden estar contraindicados en algunas ocasiones. Por ejemplo, si como resultado de la evaluación han de tomarse decisiones que repercuten directamente en la población evaluada, no es correcto

basar decisiones respecto a individuos en informaciones de grupo. En algunos casos esto resulta evidente. No podemos afirmar nada respecto de la competencia profesional de un docente, o sobre los conocimientos de un alumno, basándonos en los resultados en la evaluación de la competencia profesional de los docentes de una región, o en la evaluación de los conocimientos de todos los alumnos de un país. Ni siquiera aunque tuviésemos los resultados específicos de ese sujeto estaría justificada dicha práctica. No ya sólo por cuestiones morales de respeto a la confidencialidad, sino sobre todo porque lo que puede representar una muestra muy informativa de las conductas o conocimientos de un sujeto de una muestra, puede no serlo de las conductas o conocimientos del sujeto «per se». Esto es especialmente cierto en los diseños «matrix sampling» en los que de cada sujeto se obtiene sólo una parte de la información relevante.

Pero aunque no es tan evidente, tampoco pueden tomarse decisiones sobre un centro basándonos en los resultados de uno de sus grupos que fue seleccionado para pertenecer a una muestra nacional. Ese elemento muestral es adecuado para la evaluación nacional, y aporta una cierta cantidad de información a ese propósito. Sin embargo no es representativa del centro del que procede, ya que el dominio no coincide con el mismo. Ésta es una de las muchas tentaciones en las que pueden caer los representantes políticos o los administradores cuando se encuentran con la información que proporciona una evaluación. Son muchas las posibilidades de uso indebido de la información procedente de los muestreos, y por eso es tan importante que desde su inicio estén bien definidos los objetivos de la evaluación y que los sujetos y las instituciones evaluadas sepan de antemano cuál es el uso que se va a dar a esa información, y sobre todo, qué uso no se va a dar jamás a la misma.

Ciertamente no es este el lugar para desarrollar la teoría del muestreo, ni los distintos procedimientos. Sin embargo podemos destacar aquellos conceptos relevantes al muestreo para la evaluación, y aclarar algunos problemas que en la misma se presentan.

Nos vamos a referir por tanto a la población objetivo, a los dominios, a los estratos, a los conglomerados, a la correlación intra-clase, al muestreo de los conglomerados, al tamaño de la muestra o las ponderaciones de los datos y sus efectos, y a un caso especial en el muestro.

2. LA POBLACIÓN OBJETIVO

Cuando se planifica una evaluación, una cuestión que aparentemente está siempre clara es cuál es la población objetivo, es decir, cuál es el grupo de sujetos, o en general unidades de análisis, a las que la información recabada por la evaluación va a hacer referencia. Por lo general una primera definición permite acercarse bastante al objetivo buscado. Por ejemplo, «todos los alumnos correspondientes al curso modal de 14 años». Esta primera definición determina cuál es la **población implícitamente excluida**. Se trata de todos los alumnos, independientemente de su edad, que correspondan a cursos cuya edad modal no sea 14 años.

Sin embargo se presentan dudas que es preciso resolver. ¿Deben incluirse los alumnos que están implicados en programas complementarios de educación especial? En

ese caso debe acudirse al objetivo de la evaluación para responder a la cuestión. Por ejemplo, podría ser el objetivo del estudio la evaluación de la acción del sistema excluido el subsistema de educación especial, cuya evaluación se realizaría en fase distinta. Eso entonces nos lleva a determinar la **población explícitamente excluida**. Así, en el estudio del INCE (1998) «se determinó como población excluida los alumnos con necesidades educativas especiales tales que impidiesen contestar la prueba en igualdad de condiciones espaciales y temporales con los demás alumnos del grupo. Se instruyó a los aplicadores para que aplicasen a estos alumnos las pruebas igual que al resto de sus compañeros, y sólo *a posteriori* deberían identificar dichas pruebas para su exclusión».

3. LOS DOMINIOS

Un dominio se refiere a una subdivisión de la población para la que se han planeado, designado y seleccionado diferentes muestras, con el objeto de proporcionar resultados específicos para la misma con un conocido margen de error. (Kish, 1987). Esto quiere decir que los resultados se proporcionarán con una precisión determinada para cada dominio, y que el resultado global procederá de una combinación adecuada de los resultados de todos los dominios. Lo habitual es que los dominios coincidan con ciertas unidades político-administrativas, tales como regiones, comunidades, provincias, etc.

Si se decide la utilización de dominios, entonces el muestreo en la práctica se realiza como si cada uno de los mismos fuese una población de la que hay que extraer una muestra representativa. Naturalmente tiene que ocurrir que la muestra agregada de todos los dominios a su vez cumpla los requisitos de representatividad establecidos para la población total.

4. LOS ESTRATOS

Es conocido que cuando se clasifica a una población según ciertas características demográficas, los grupos así formados tienden a ser más homogéneos respecto a otras variables de interés que el conjunto de la población. Así por ejemplo, la varianza en altura de los sujetos de alto nivel socio-económico familiar es menor que la varianza en altura de la población en general. Los grupos así formados se denominan estratos, y su utilización permite obtener mayor cantidad de información de la muestra utilizando menos sujetos. A esto se le llama efecto diseño, y consiste en la disminución o el aumento del tamaño de la muestra respecto al muestreo aleatorio simple debido al tipo de muestreo utilizado.

La utilización de estratos es muy recomendable en la evaluación, aunque también es posible que un numero excesivo de estratos haga la extracción de la muestra excesivamente compleja, compensando negativamente las ventajas obtenidas por la disminución de la varianza. En teoría los estratos dividen horizontalmente a los dominios. Es posible incluso que cada dominio tenga una estratificación distinta, reflejando los intereses diversos de los administradores. Pero también es posible que

387

una cierta combinación de estratos se convierta en un subdominio. Por ejemplo, es frecuente que los administradores deseen conocer con un nivel predeterminado de precisión los rendimientos en los centros públicos y en los centros privados en una región determinada. En ese caso los alumnos de centros públicos y de centros privados de dicha región se convierten en subdominios de los que habrá que extraer muestras representativas.

¿Cuántos estratos pueden definirse en un procedimiento de muestreo? La tentación vuelve a ser la de definir el mayor número posible. Sin embargo, como regla práctica, diremos que el número máximo de estratos definidos no debe ser mayor que la mitad del número de unidades primarias de muestreo. El sentido de esta regla consiste en que debe ser posible calcular el error muestral, y para ello es necesario al menos dos de estas unidades por cada estrato.

Las variables de estratificación más habituales son la clasificación administrativa (provincia, estado, región, comunidad autónoma), ámbito rural o urbano, nivel socioeconómico del distrito escolar o del entorno de la escuela, naturaleza pública o privada del centro, tamaño del centro, etc.

5. LOS CONGLOMERADOS Y LA CORRELACIÓN INTRA-CLASE

Una particularidad que el sistema educativo comparte con otras organizaciones sociales es que los sujetos forman grupos que comparten tiempo y actividades. De hecho acceder a un solo individuo es prácticamente tan costoso como acceder a todo el grupo primario, en este caso el aula. Lógicamente conviene entonces recabar información no de un solo sujeto, sino de los sujetos de toda el aula. Este grupo primario es un conglomerado. El hecho de que todos los individuos del conglomerado compartan tiempo, y actividades, además de que probablemente proceden del mismo entorno social, etc., hace que estos sujetos sean entre sí más homogéneos que con relación a los demás individuos de la población. Pero el efecto de esta mayor homogeneidad en los conglomerados tiene un efecto contrario al que se producía en los estratos.

Supongamos que para obtener una muestra representativa de una población necesitamos una muestra de 300 sujetos por muestreo aleatorio simple. Si el investigador accede a centros en los que hay 30 sujetos por aula, podría parecer que visitando 10 centros obtendría toda la información necesaria.

Pero imaginemos por un momento que la acción del sistema educativo es tan fuerte que todos los alumnos de un mismo centro fuesen exactamente iguales en las variables de interés, es decir, que la varianza dentro de los centros fuese cero, y que todas las diferencias en la población fuesen debidas a la diferente influencia de cada centro. En ese caso, la información que proporciona un centro se obtiene con un solo sujeto, y el resto de los individuos proporcionaría información redundante. Esto quiere decir que cuando el investigador evalúa a los treinta sujetos de un centro la información obtenida no es mayor que la que proporciona uno sólo. Por tanto después de visitar 10 centros, y aunque haya evaluado a sus trescientos sujetos, tiene información correspondiente sólo a 10 de los 300 individuos necesarios para que su muestra sea representativa.

388

En el otro extremo, si los sujetos fuesen completamente heterogéneos dentro de las aulas, y los centros fuesen completamente equivalentes unos y otros, es decir, si toda la varianza en la población fuese debida sólo a las diferencias intra-aula, bastaría con recoger la información de esos 10 centros para completar la muestra.

El índice que mide la mayor o menor homogeneidad de los sujetos en los conglomerados es la correlación intra-clase. De hecho la correlación intra-clase es una medida de la proporción de la varianza total que es debida a las diferencias dentro de los conglomerados.

En el muestreo por conglomerados el efecto diseño tiene un efecto negativo. Esto quiere decir que cuanto mayor sea la correlación intra-clase, mayor deberá ser el tamaño de la muestra final respecto de lo que hubiera sido si se tratase de muestreo aleatorio simple.

6. EL MUESTREO DE LOS CONGLOMERADOS

Que la muestra sea representativa supone que todos los sujetos de la población objetivo, o en su caso del dominio, tengan la misma probabilidad de pertenecer a la muestra. Cuando las unidades primarias de muestreo son los conglomerados, es decir, las aulas, si todas las aulas tienen la misma probabilidad de formar parte de la muestra, entonces los sujetos que están en aulas más pequeñas tienen mayor probabilidad de formar parte de la muestra que la que en puridad les corresponde. Esto quiere decir entonces que las aulas no deben tener la misma probabilidad de formar parte de la muestra. Las aulas pequeñas deben tener menor probabilidad y las grandes mayor. Es decir, el muestreo debe realizarse con probabilidad proporcional al tamaño del aula.

7. EL TAMAÑO DE LA MUESTRA

Es una creencia muy generalizada entre los no especialistas que lo que determina esencialmente el tamaño de la muestra es el tamaño de la población a que esa muestra tiene que representar. Según esto, para una población de 100.000 sujetos, correspondería una muestra de doble tamaño que para una población de 50.000 individuos. Sin embargo, no es así. Para un nivel de confianza dado, los parámetros que más afectan al tamaño de la muestra son el error muestral máximo que se está dispuesto a asumir, la desviación típica de la variable medida, y la correlación intra-clase en el caso del muestreo por conglomerados. En la gráfica 1 vemos la relación que existe entre el tamaño de la desviación típica de la población y el tamaño necesario de la muestra para un error muestral máximo de 1,5 puntos a un nivel de confianza del 95%, en dos poblaciones distintas, una con 12.000 sujetos, y otra mucho más grande de 450.000. Como vemos para una sigma de 9,5 puntos por ejemplo, el tamaño de la muestra que garantiza un error no mayor de 1,5 es de 158 sujetos en una población de 12.000 sujetos, y de 160 en una población de 450.000.

Por tanto, garantizando un cierto valor de error muestral máximo el tamaño de la población no va a afectar de forma sustancial al tamaño de la muestra.

389

Por esta razón las diferencias en el tamaño de las muestras utilizadas en los distintos dominios definidos en una evaluación, provincia, región, comunidad autónoma, no vienen determinadas por el tamaño relativo de la población de cada una de ellos, sino por el error muestral máximo que se esté dispuesto a asumir en cada una de ellos.

Por distintos motivos, especialmente cuando los dominios definidos coinciden con unidades político-administrativas distintas, suele ocurrir que se definen distintos márgenes de error para cada uno. Esto depende del tipo de información en que la burocracia de cada una de estas unidades tiene mayor interés, del presupuesto disponible y de la influencia que pueda tener sobre el órgano que gobierna la planificación y realización de la evaluación.

Para la determinación del tamaño muestral es necesario no sólo el error muestral máximo, sino también el valor de la correlación intra-clase. La estimación de los valores de este parámetro pueden obtenerse, bien por referencias previas, o bien por la aplicación de una prueba piloto. A modo de ejemplo, en la evaluación de la secundaria realizada por el INCE en 1997, la aplicación de la prueba piloto permitió obtener una estimación de los valores de la correlación intra-clase para los grupos de 14 y de 16 años, siendo esos valores de 0,13 y 0,22 respectivamente.

La práctica habitual consiste en fijar un determinado nivel de confianza, y realizar una primera estimación del muestreo que sería necesario en caso de Muestreo Aleatorio Simple (M.A.S.). A continuación se calcula la corrección necesaria por el efecto diseño debido al Muestreo Aleatorio por Conglomerados, según se explica en el apartado siguiente.

8. PROCEDIMIENTO DE MUESTREO

El procedimiento de selección de la muestra se basa en el supuesto de equiprobabilidad de selección de todos los sujetos. Si no atendiésemos a la estructura natural de los datos, es decir, agrupados en aulas, a su vez agrupadas en colegios, determinaríamos el máximo error muestral que estaríamos dispuestos a admitir para cierto nivel de confianza, y a partir de estos parámetros calcularíamos el tamaño de la muestra mediante la expresión

$$n = \frac{N\frac{k^2}{e^2}}{N + \frac{k^2}{e^2}}$$

donde N es el tamaño de la población de referencia, k es la puntuación típica correspondiente al nivel de confianza seleccionado, y e es el error muestral expresado en desviaciones típicas. Esta expresión contiene ya la corrección por poblaciones finitas.

Esta fórmula es de aplicación para cada uno de los dominios.

330

La muestra total será la agregación de las muestras de todos los dominios.

Dado que es de interés proporcionar información acerca de algunos estratos de la población, se dividirá a ésta en varios estratos. Es el caso por ejemplo de las provincias, de líneas curriculares, de la titularidad de los centros, o cualesquiera otros. Por tanto, para proporcionar datos acerca de cada uno de los estratos, el número de sujetos que se determina para un dominio se reparte proporcionalmente entre los distintos estratos en que se divide su población, de manera proporcional al tamaño de estos.

Pero los sujetos están agrupados, dentro de cada estrato, en conglomerados, constituidos por las aulas, y estas aglomeradas en colegios.

Por esta razón cuando los sujetos forman conglomerados, el tamaño de la muestra en comparación con el muestreo aleatorio simple, debe aumentar en función del tamaño del conglomerado y de la autocorrelación. A ese incremento en el tamaño se le llama efecto diseño, y para el caso que nos ocupa, viene dado por $F=1+(B-1)\rho$, dónde B es el tamaño del conglomerado, y ρ es la autocorrelación.

Por tanto el tamaño muestral definitivo será $n'=Fn$. Como los sujetos están agrupados de B en B, se seleccionarán por tanto n'/B aulas distintas.

Para ilustrar este punto tomaremos un caso de la mencionada evaluación de la secundaria (INCE 1998). Para la comunidad autónoma de La Rioja los cálculos serían los siguientes. En el grupo de 14 años hay 3.541 alumnos en la población. Si queremos que el nivel de confianza sea del 95,5%, y el error muestral máximo sea de 0,12 sigmas, obtenemos un tamaño para muestreo aleatorio simple de 257,57 alumnos. Como el valor estimado de rho para los 14 años era de 0,13, el efecto diseño era igual a 4,12. Por tanto en un muestreo por conglomerados le correspondería un tamaño de muestra de 1.061 alumnos.

Si comparamos ese resultado con el tamaño de muestra que correspondería a una comunidad proporcionalmente mucho más grande como Cataluña, podemos comprobar lo poco que el tamaño de la población afecta al tamaño de la muestra. Así, manteniendo los demás parámetros iguales, es decir, el nivel de confianza y el error muestral máximo, dado que Cataluña tiene 73.293 alumnos de 14 años, en un muestreo aleatorio simple le correspondería un total de 267,72 alumnos (véase la poca diferencia con los 257,57 que correspondían a La Rioja). Aplicando la corrección del efecto diseño, con el mismo rho, tenemos que en el muestreo por conglomerados le correspondería un total de 1.103 alumnos de 14 años. (En la realidad la comunidad autónoma de Cataluña decidió extraer la muestra de 14 y la de 16 años con menor error que otras comunidades autónomas, lo que explica el mayor tamaño de su muestra de 14 años. Lo mismo se aplica a la muestra de 16 años).

Un caso muy habitual en la evaluación de resultados educativos es el muestreo en dos etapas. En primer lugar se seleccionan n'/B centros, con probabilidad proporcional a su tamaño. Y dentro de cada centro, se selecciona en la siguiente etapa un aula completa con probabilidad también proporcional a su tamaño.

Este procedimiento debe repetirse para cada población objetivo distinta. No importa que dos poblaciones objetivo estén localizadas en las mismas unidades primarias de muestreo, en nuestro caso colegios. Se trata de mantener el supuesto de equiprobabilidad. Si la selección de un centro se hace con probabilidad proporcional al número de

391

sujetos de 14 años que tiene, si seleccionásemos también a otra muestra de sujetos de 16 años ocurriría que en esta población no todos los sujetos tendrían la misma probabilidad de ser incluidos en la muestra. El procedimiento de muestreo debe por tanto repetirse para cada población objetivo distinta.

Con relación a la muestra global para toda la población evaluada, se constituye por la agregación de las muestras de los dominios. Esta forma de muestreo implica un muestreo no proporcional, ya que el número de sujetos de cada dominio puede no ser proporcional a N_h/N. Esto lleva de forma inmediata a la necesidad de calcular la media global con ciertos coeficientes de ponderación, que hacen que cada dominio esté representado en la media total de forma proporcional a su tamaño relativo de la población.

Las pequeñas diferencias entre las muestras planificadas y las muestras obtenidas dentro de los subestratos de cada dominio también aconsejan utilizar ciertas ponderaciones para lograr que en la media de cada dominio estén respresentadas proporcionalmente las poblaciones de cada subestrato.

Para más información puede consultarse Blalock (1972) y Kalton (1983).

9. DETERMINACIÓN DE LAS PONDERACIONES

Como consecuencia del muestreo por dominios, aquéllos que son más pequeños quedan sobrerepresentados, mientras que los más grandes quedan subrepresentados.

Para corregir ese efecto es preciso que a cada dominio se le asigne un peso de forma que su muestra tenga la importancia que le corresponde en el total. Para que las estimaciones sean correctas ese peso tiene que ser tal que su media total sea igual a 1.

Hay otra fuente de desviaciones que hacen necesario el uso de otras ponderaciones. Entre la muestra planificada y la muestra prevista existen ciertas discrepancias producidas por las contingencias propias del proceso de toma de datos. Recordemos que se trata de un muestreo por conglomerados. Sólo aproximadamente puede anticiparse el número de sujetos en cada conglomerado. Dentro de cada dominio, se definen cuáles son los estratos de interés. Dado que en última instancia el tamaño final de la muestra depende de los verdaderos tamaños de los conglomerados, se dan ciertas pequeñas discrepancias entre el tamaño planificado y el tamaño obtenido. Como consecuencia los elementos de cada estrato, dentro de un mismo dominio, deben ser también ponderados para que ninguno de ellos tenga un mayor peso que aquél que le corresponde. Estos pesos intradominios, que deben ser utilizados siempre que se realicen comparaciones entre distintos estratos de un mismo dominio, deben cumplir también la condición señalada anteriormente, por la cual la media de estos pesos debe ser igual a uno.

Ciertamente cuando se realizan las estimaciones de las medias globales, debe tenerse en cuenta tanto las correcciones debidas a las desviaciones de los estratos dentro de cada dominio, como las ponderaciones atribuibles a cada dominio respecto del conjunto.

En general, para calcular el peso que corresponde a un cierto estrato *i*, procederemos de la siguiente manera.

Si llamamos
a_i =tamaño del estrato *i* en la muestra
A_i =tamaño del estrato *i* en la población.
n =tamaño de toda la muestra
N= tamaño de la población.
Entonces en un dominio con k estratos, la media total del dominio es

$$\bar{x}_t = \frac{\sum_i^k A_i \frac{\sum_j^{a_i} x_{ji}}{a_i}}{N} = \sum_i^k \frac{A_i}{a_i N} \sum_j^{a_i} x_{ji} \qquad \text{Luego,}$$

$$p_i = \frac{A_i}{a_i N}$$

Aquí la media se obtiene como la suma ponderada de todas las puntuaciones. Pero, nos conviene un peso para cada puntuación, de forma que la media, no la suma, de todas las puntuaciones nos permita obtener la media total. Así que para dividir por *n* todo, tenemos que multiplicar por *n* el peso también.

De esta forma obtenemos un peso tal que multiplicando el cociente del tamaño del estrato en la muestra respecto a la muestra, éste es proporcional al mismo cociente en la población.

$$\frac{a_i}{A_i} p_i = \frac{n}{N}$$

$$p_i = \frac{nA_i}{Na_i}$$

Estos pesos tienen la propiedad de que su media es 1. Si las ponderaciones no tienen como media 1, entonces los niveles de significación de que informan los procedimientos estadísticos no son válidos.

10. LA PONDERACIÓN Y LA MEDIA DE LOS ESTRATOS

Cuando se calculan las medias de los estratos individualmente utilizando los pesos, las medias de los estratos no varían.

$$\bar{x}_i = \frac{\sum_j^{a_i} p_i x_{ji}}{\sum_j^{a_i} p_i} = \frac{p_i \sum_j^{a_i} x_{ji}}{a_i p_i} = \frac{\sum_j^{a_i} x_{ji}}{a_i}$$

En cuanto a la media total usando los pesos,

y como $a_i p_i = \frac{A_i}{N} n$

$$\bar{x}_t = \frac{\sum_i^k \sum_j^{a_i} p_i x_{ji}}{\sum_i^k \sum_j^{a_i} p_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\sum_i^k a_i p_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\sum_i^k \frac{A_i}{N} n} =$$

$$= \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\frac{n}{N} \sum_i^k A_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\frac{n}{N} N} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{n}$$

11. EFECTO DE LA PONDERACIÓN SOBRE LA VARIANZA

En los procedimientos estadísticos utilizados habitualmente la variable de ponderación se utiliza para «replicar» los casos. De esta manera tenemos para un estrato.

$$\sigma^2_{p_i x_i} = \frac{\sum (x_{ji} - \bar{x}_i)^2 p_i}{\sum_j^{a_i} p_i - 1} = \frac{p_i \sum_j^{a_i} (x_{ji} - \bar{x}_i)^2}{a_i p_i - 1} = \frac{\sum_j^{a_i} (x_{ji} - \bar{x}_i)^2}{a_i - \frac{1}{p_i}}$$

Esta es la estimación de la varianza dentro del subestrato *i*. Así cuando $p_i > 1$, la varianza de las puntuaciones ponderadas es inferior a la varianza de las puntuaciones sin ponderar. Sin embargo esa diferencia es muy pequeña cuando los valores de la ponderación son los que proceden de los pequeños errores de muestreo en los subestratos.

12. ALGUNOS CASOS ESPECIALES

Por último nos vamos a referir a algunos casos especiales que surgen en la evaluación. Un problema muy interesante en el muestreo aparece cuando tratamos de evaluar sujetos que pertenecen a un grupo muy minoritario, como por ejemplo aquéllos que fueron definidos como parte de la población excluida. En ocasiones es difícil

determinar cuántos son los sujetos que están en esas condiciones en la población de referencia. Es el mismo caso cuando tratamos de determinar por ejemplo cuál es, en una población, la proporción de sujetos que presentan algunas características especiales, tales como una alta capacidad matemática, o especial habilidad para la música, o tienen problemas de aprendizaje, etc.

La determinación de la proporción en la población, es decir, el parámetro π por los medios tradicionales puede ser muy costoso en términos de tamaños de la muestra o de error de estimación. El tamaño de la muestra viene dado por

$$n = \sigma_p^2 \frac{z_{\alpha/2}^2}{E^2}$$

como puede verse el tamaño de la muestra depende radicalmente del inverso de la precisión que deseamos obtener. Si queremos tener una precisión de 1 por mil al 95% de confianza, el tamaño de la muestra será 4000000 de veces la varianza de p. Para un fenómeno que ocurra alrededor de 5 casos cada mil sujetos de la población, la muestra necesaria para estimar ese parámetro es de cerca de 20.000 sujetos. Esto hace inviable la mayoría de estudios de este tipo.

Existen sin embargo alternativas muy interesantes. Por ejemplo los procedimientos de captura-recaptura, muestreo inverso para proporciones, muestreo de redes. Mencionaremos el segundo, por ser el más adecuado al tema tratado. Los otros dos pueden consultarse en Som (1996).

En el muestreo inverso se procede a seleccionar sujetos para la muestra hasta que se obtiene un número n' determinado de casos buscados. Si n es el tamaño final de la muestra, la estimación del parámetro buscado es

$$p = \frac{n'-1}{n-1}$$

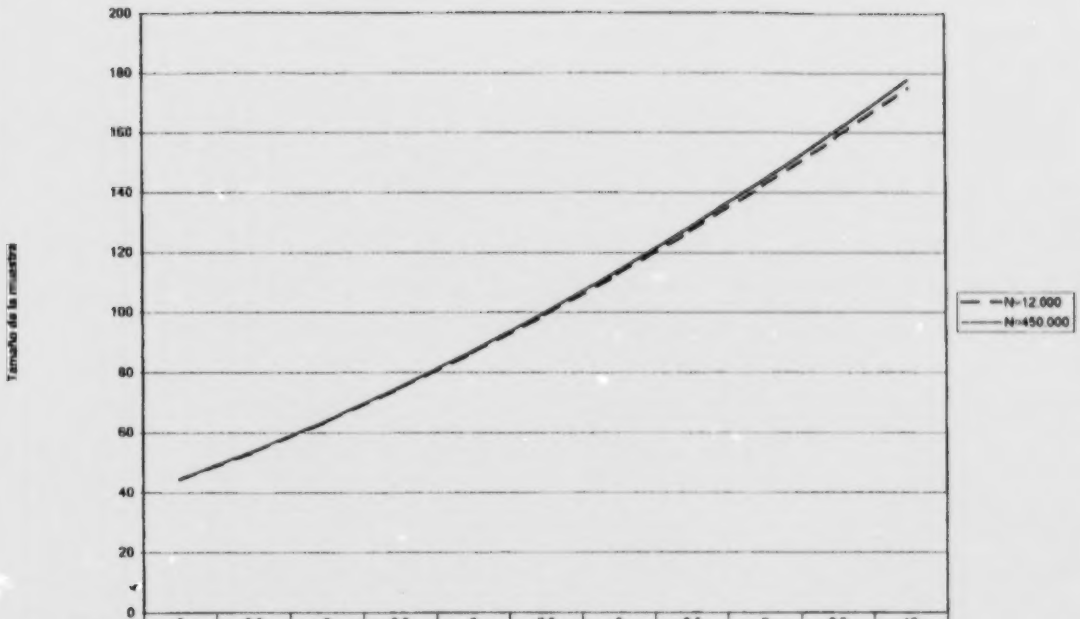
la varianza del estimador viene dada por $s_p^2 = \frac{p(1-p)}{n-2}(1-\frac{n-1}{N})$, siendo N el tamaño de la población.

Existen estimadores también para variables continuas en los mismos casos.

13. BIBLIOGRAFÍA

Blalock, H.M. (1972). Social Statistics, New York: McGraw-Hill.
INCE (1998) 2. Los Resultados escolares. Ministerio de Educación y Cultura. Madrid.
Kalton, G. (1983). Introduction to Survey Sampling, Sage, London.
Kish, L. (1987). On the future of survey sampling. En N.K. Namboordi (Ed.) Survey sampling and measurement. Academic Press. N.Y.
Som, R.K. (1996). Practical Sampling Techniques. Marcel Dekker, Inc. N.Y.

395



	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
—N=12.000	44	54	64	75	88	90	113	127	142	158	175
——N=450.000	44	54	64	75	87	100	114	128	144	160	178

Sigma

Gráfico 1
Relación entre desviación típica y tamaño de la muestra

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 405-416

LA EVALUACIÓN Y MEJORA DE LA ENSEÑANZA EN LA UNIVERSIDAD: OTRA PERSPECTIVA

Tomás Escudero Escorza
Universidad de Zaragoza

RESUMEN

Tomando la perspectiva del responsable de la gestión académica universitaria, se analizan las características de la evaluación de la enseñanza en la universidad y los ámbitos preferentes de actuación para mejorarla. La ubicación de los estudiantes, la formación y la calidad del profesorado, los diseños curriculares y la metodología didáctica y sus recursos son analizados de manera expresa, destacando ciertos elementos y ofreciendo algunas sugerencias para mejorar nuestra enseñanza universitaria.

ABSTRACT

Taking the perspective of people in charge of academic affairs, characteristics of teaching evaluation at the university are analyzed and considerad sume important spaces for improvement. Student selection of degrees, training and quality of teachers, curricula designs and teaching methodology and resources are taken as main elements for improvement after evaluation of teaching in our universities.

Participar en un número monográfico como el presente, en unión de tantas plumas cualificadas es, sin duda, un honor, pero tiene el riesgo de que las reflexiones que se ofrezcan sean reiterativas y reproduzcan algo ya aportado por otros colegas.

Para minimizar el riesgo apuntado, y toda vez que somos varios los autores que incidimos en el tema de la evaluación de la calidad de la enseñanza universitaria, en este artículo pretendo ofrecer la visión del responsable universitario y, más en concre-

to, la visión desde la que se ve obligado a trabajar alguien que, como es mi caso, ha tenido la suerte y la satisfacción de ser Vicerrector de Evaluación y Mejora de la Enseñanza, durante los últimos cuatro años, en una importante Universidad española como lo es la de Zaragoza.

Como es lógico, la perspectiva que ofrezco nunca puede ser considerada la de un gestor universitario en sentido convencional, pues resulta imposible desligarla de mi condición de profesor y de estudioso del tema sobre el que estamos reflexionando. Sin embargo, sí que intento enfatizar aquello más específico de las preocupaciones concretas de un responsable institucional que tiene la obligación de tomar y promover decisiones encaminadas a mejorar la enseñanza de su universidad. Con este planteamiento, espero y deseo que mi aportación complemente y enriquezca otras que aparecen en este número monográfico de RIE, desde diferentes perspectivas académicas y personales.

Este punto de partida y la propia limitación de espacio que se me sugiere, me obliga a no entrar en análisis de tipo general sobre modelización y procedimientos de evaluación de los programas y de la enseñanza universitaria, por lo que me centraré en el análisis de ciertos aspectos que, entiendo, son cruciales para valorar y mejorar la enseñanza de nuestras universidades.

1. EL OBJETIVO DE LA EVALUACIÓN

Los profesionales de la evaluación estamos continuamente repitiendo que esta no es un fin en sí misma, que es un escalón, una plataforma intermedia, que facilita y perfecciona las decisiones y acciones de planificación, gestión, desarrollo y mejora. Sin embargo, uno tiene la impresión de que dicho mensaje no acaba de calar suficientemente entre los implicados y afectados por la evaluación. Desgraciadamente, hoy hacemos mucha evaluación que parece cerrarse en sí misma, que temporalmente produce cierto impacto informativo en unas u otras audiencias, pero que no tiene consecuencias palpables sobre la calidad de nuestra enseñanza. A pesar de tanta reflexión teórica al respecto, paradójicamente, en el terreno de la práctica, evaluación, calidad y mejora no suelen hacerse a menudo confluir, como sería su obligación.

Es muy cierto que de esta paradoja no suelen ser responsables principales los evaluadores, sino los que tienen la capacidad de decisión académica, gestora, legislativa o presupuestaria, pero no es menos cierto que se adolece de cierta tibieza en la defensa de este principio básico. Y en esto hay que ser radical. O la evaluación es útil o no tiene sentido realizarla; tiene que ser un instrumento para la acción y no un mero mecanismo de justificación o para tranquilizar conciencias. Los evaluadores debemos ser beligerantes en este sentido.

Lo dicho nos obliga a pensar que el diagnostico riguroso y sensato de la realidad no es el final de nuestra acción evaluadora, tenemos que ofrecer, además, sugerencias alternativas de acción, con sus posibles riesgos, con sus posibles ventajas. En definitiva, estamos obligados a defender el estatus científico y el marco investigador propio del evaluador institucional (Fishman 1988 y 1991; Escudero, 2000).

2. A VUELTAS CON LA CALIDAD DE LA ENSEÑANZA Y SUS POSIBLES INDICADORES

Sin entrar en profundidad en el siempre interesante debate sobre el concepto de calidad de la enseñanza, sí que es importante señalar algunos elementos de referencia básicos de dicha calidad que, en definitiva, van a ser el objeto central de nuestra evaluación.

Los responsables universitarios y la sociedad en general tienen muy claro que el *indicador directo* de la *calidad de la enseñanza* son sus *resultados*, esto es, el nivel formativo alcanzado por los estudiantes. Pero esta idea generalizada, que además es correcta en su base, tenemos la obligación de matizarla y contextualizarla convenientemente, para que sea interpretada de manera cabal y rigurosa.

Al margen de las graves dificultades para contar con escalas estandarizadas de medición de los niveles formativos alcanzados por los estudiantes en distintas carreras, cursos, especialidades, disciplinas,..., que deberíamos utilizar como instrumentos para la homologación y comparación, los resultados directos que podemos obtener expresan la *calidad de los estudiantes* y, en su caso, de los titulados y no la de la enseñanza, pues tales resultados están condicionados por las capacidades, el interés y el esfuerzo de los alumnos. Por desgracia, en el mejor de los casos, lo que podemos medir es la interacción entre la calidad de la enseñanza y la de los alumnos.

Por confundir los dos conceptos anteriores, calidad de enseñanza y de estudiantes, no son pocos los ciudadanos, e incluso académicos y responsables académicos, que asimilan incorrectamente la selectividad universitaria con calidad de la docencia. La selectividad es, sin duda, un elemento facilitador de la docencia universitaria, pero en ningún caso un indicador de su calidad. Lo que sí puede ser es un factor determinante en la calidad de los estudiantes. Lamentablemente, al cobijo de esta confusión de ideas, algunas instituciones han construido una buena imagen social de calidad docente con prácticas pedagógicas criticables y programas académicos mediocres, fomentando el rigor académico y social en la selección de sus estudiantes.

Para complicar todavía un poco más el panorama que se obseva y desde el que se actúa, aparecen las calificaciones escolares como instumento sobre el que se construyen los populares conceptos de éxito o fracaso escolar. Estas calificaciones, como medida de los resultados de la enseñanza, no sólo están sujetas a todos los condicionantes planteados anteriormente, sino que, además, no son un indicador homogéneo, pues dependen del criterio y rigor del profesor a la hora de diseñar la enseñanza y valorar el rendimiento de los estudiantes. En tal sentido, las calificaciones escolares son un indicador muy débil de la calidad de nuestra enseñanza, por lo que siempre deben ser complementadas con otro tipo de indicadores y criterios (De Miguel, 1999; Ferrández y otros, 1999).

El interés de las calificaciones escolares, desde la perspectiva de la evaluación, está en que sí son un buen indicador de la *coherencia interna* y de la *eficiencia institucional* en el ámbito docente, además de un elemento de diagnóstico imprescindible para tratar con las bolsas de fracaso escolar. Por añadidura, no podemos olvidar que se trata de

399

un elemento clave a la hora de construir la imagen social de una institución de educación superior.

Por lo dicho hasta aquí, entiendo que las *calificaciones escolares*, con sus pros y sus contras y siempre que se contextualicen debidamente, son un elemento de análisis importante para el diagnostico de la calidad intitucional y para ubicar el fracaso escolar (Escudero, 1999). No podemos olvidar que uno de nuestros grandes retos es el de atajar el elevado fracaso escolar que se produce en algunos cursos de alguna de nuestras titulaciones. Para ello debemos ubicarlo con precisión, pues el panoramay la etiología son muy diversos. No nos basta con detectar una elevada tasa de fracaso en tal o cual titulación, en tal o cual curso, pues sabemos muy bien (Escudero y otros, 1999), que el bloqueo puede limitarse a una o dos asignaturas y no al resto de la titulación. El evaluador está obligado a indicar el espacio preciso en el que se debe intervenir.

Pero incluso trabajando a un nivel suficientemente desagregado de la información, que nos permita diagnostigar con mucha precisión el lugar del fracaso, al analizar los resultados de la enseñanza, las soluciones y pautas de intervención vamos a tener que buscarlas a través de *indicadores indirectos* de la calidad de la enseñanza, relacionados con las entradas y el proceso docente (Segers y Dochy, 1996). En los siguientes apartados centramos la reflexión en algunos de los más críticos indicadores indirectos de la calidad.

En paralelo con los indicadores anteriores, las instituciones de educación superior utilizan de manera generalizada *indicadores de impacto* o niveles de satisfacción de los afectados por la enseñanza, como indicadores de calidad de la misma. La experiencia al respecto me dice que este tipo de evaluación es muy importante, pero que, en la práctica, es sistemáticamente infrautilizada como instrumento de mejora de la enseñanza, bien porque se trabaja con niveles de agregación de opiniones demasiado globales, bien porque no se analizan, o no se tienen en cuenta los resultados y lo que sugieren (Escudero y otros, 1997 y 2000; Escudero, 1999a). La opinión de los estudiantes y de los profesores, sin duda, es uno de los principales caminos que tenemos los evaluadores y, consiguientemente, las universidades para encontrar y diseñar acciones de cambio, de innovación y de perfeccionamiento. Tampoco podemos olvidar que en este proceso, resulta crítico delimitar bien los espacios de responsabilidad pues, como bien sabemos, las universidades son organizaciones de organizaciones poco cohesionadas, con múltiples y diversos centros de autonomía y de responsabilidad. Los evaluadores tenemos que decirle a la institución dónde y en qué sentido tiene que intervenir, pero también quien es el que tiene el deber y la responsabilidad de actuar.

3. ÁMBITOS DE EVALUACIÓN PARA LA MEJORA

Una vez contextualizados el problema y el marco de actuación, es el momento de plantearse el proceso de evaluación de la enseñanza que, desde una perspectiva ortodoxa, nos llevaría al apoyo en enfoques reconocidos y clásicos en el campo de la evaluación de programas (Municio, 1992; Pérez Juste, 1995; Gredler, 1996). Sin embargo, de acuerdo con la perspectiva desde la que he planteado este trabajo y respetando sus objetivos, voy a centrarme en la evaluación de aquellos aspectos, facetas o ámbitos que se sabe que inciden de manera decisiva en la calidad de la enseñanza universitaria y

sobre los que tienen una responsabilidad directa de transformación los gestores y responsable académicos de la institución. En concreto, quiero destacar cuatro grandes ámbitos de mejora de nuestra enseñanza, a los que resulta obligado dirigir la evaluación y la intervención en los momentos actuales en nuestras universidades:

a) La ubicación de los alumnos en los distintos estudios
b) La formación y la calidad del profesorado
c) Los diseños curriculares
d) La docencia propiamente dicha, su metodología y sus recursos.

3.1. Ubicación de los alumnos

Aunque ya he señalado el peso relativo que tiene la selección de los alumnos en la determinación de la calidad de la enseñanza, es cierto que una buena ubicación de los mismos en los diferentes estudios facilita el desarrollo del proceso docente y la consecución de más y mejores logros por su parte.

Con la legislación actual, el margen de maniobra de las universidades en este terreno es limitado, centrándose principalmente en el establecimiento de los límites de acceso, allí donde se produzca exceso de demanda. Ahora bien, ¿se está haciendo todo lo posible en cuanto a la orientación en la elección de estudios, que no es un aspecto cerrado por la legislación?

La impresión es que nuestras universidades, en colaboración con los centros de secundaria, están dando mucha información, quizás excesiva, sobre el acceso a la educación superior, pero con resultados escasos para el esfuerzo aportado. Es necesario, por tanto, evaluar estos procesos, detectar los puntos de ineficiencia y sugerir nuevas alternativas en este terreno.

La orientación vocacional y profesional, aunque se ubica de manera más directa en los centros de secundaria, no puede estar al margen de las preocupaciones de los diferentes centros universitarios, que deben colaborar, y proporcionar todo tipo de información, con los gabinetes de orientación. Las universidades tienen que establecer unas redes de colaboración en este sentido y evaluar y mejorar de manera sistemática su funcionamiento.

Otro elemento a considerar por los evaluadores es el de los criterios para el establecimiento del límite de plazas en las diversas titulaciones, no siempre en apoyo de una mejor calidad y organización docente. Con frecuencia aparecen criterios fruto de presiones e intereses espúreos. Aprovechando el descenso en la demanda, las universidades deben evaluar y revisar los viejos criterios de establecimiento de limitación de plazas de estudio, ajustándolos a razones mejor fundamentadas desde los puntos de vista social y académico.

3.2. Formación y calidad del profesorado

Es casi una obviedad decir que un elemento clave de una buena docencia es el profesorado. Tener buenos profesores no es condición suficiente para tener una buena

docencia, e incluso una buena enseñanza, pero es obligado. Por lo tanto, la calidad de los profesores debe ser preocupación permanente de la evaluación institucional.

Aunque no es momento de profundizar en el tema, sabemos que no es fácil definir de manera concisa qué es un buen profesor; hay muchos posibles elementos y aproximaciones. Lógicamente, tampoco resulta fácil medirlo. Sin embargo, sí que tenemos claro que se necesitan cierta características básicas para que la calidad sea posible. Desgraciadamente, algunas de estas cualidades necesarias no son fáciles de proporcionar o, más propiamente, fácilmente mejorables. A veces la intervención en este terreno tiene necesariamente unos efectos limitados y, en ciertos casos, el objetivo no podrá ser la obtención de buenos profesores en sentido estricto, sino el de evitar malas prácticas docentes.

Buena parte del problema se centra en la *selección del profesorado*, donde la capacidad de incidencia del conjunto de una universidad es mucha, aunque existan condicionantes y elementos de distorsión. Personalmente, tengo pocas dudas de que si la selección del profesorado fuera adecuada, se evitarían buena parte de los problemas docentes que más tarde aparecen. La experiencia nos dice que cuando se selecciona a un profesor con muchas dudas sobre su futura conducta docente, es alta la probabilidad de que luego plantee problemas en este sentido.

En primer lugar hay que asegurar que los profesores tengan una sólida formación en su área de conocimiento, mucho más allá de la especialización de la tesis doctoral, con capacidad y solvencia académica, cuidando, probablemente más que en el pasado cercano, los perfiles docentes en las convocatorias de plazas. La evaluación institucional de la enseñanza universitaria no puede estar al margen de este problema, y tampoco de los procesos de *adaptación y reconversión* de profesorado, como obligado vehículo para armonizar los potenciales docentes departamentales, tras necesarios cambios estratégicos en las ofertas y demandas de las universidades.

En materia de *formación pedagógica*, inicial y permanente, del profesorado universitario el camino por recorrer es mucho, pero ya se tiene un bagaje amplio de experiencias positivas en los últimos años. Ahora, la evaluación no sólo debe plantearse como vehículo para detectar y resolver carencias pedagógicas, sino también para revisar críticamente las actividades de formación que se están llevando a cabo y promover y desarrollar nuevas fórmulas que mejoren y complementen las actuales.

La experiencia nos dice que debemos resolver problemas sobre la integración de la formación pedagógica en el curriculum personal y en los méritos del profesorado, que es necesario experimentar sobre los diseños de la formación en todos los ámbitos, contenidos, extensión, organización modular, temporalización..., y que es fundamental potenciar un banco suficiente, en cantidad y variedad, de expertos en didáctica universitaria. Valorar las deficiencias y necesidades en este sentido es otro reto de la evaluación institucional.

3.3. Planes de estudio y diseños curriculares

Parece evidente para todos que buena parte de los problemas de nuestra enseñanza universitaria, especialmente aquellos sobre los que los académicos tenemos más capaci-

dad de incidencia, se centran en el terreno curricular, esto es, en el diseño de los planes de estudio, de sus asignaturas, de la jerarquización y coordinación de las mismas, de los contenidos y niveles, etc. y, en general, en todo lo relacionado con la ordenación docente.

No creo equivocarme al señalar que la evaluación de la enseñanza universitaria debe indagar con mucha atención si los planes docentes están sobrecargados de asignaturas, si estas están sobrecargadas de contenidos y si estos están planteados a niveles de dificultad razonables. Además, deben analizarse posibles problemas de coordinación entre asignaturas, de jerarquización de conceptos y de contenidos de dudosa utilidad tanto académica como práctica. La escasa atención a los problemas citados es causa de exigencias disímiles, con consecuencias indeseables en el terreno de la motivación.

Para atajar anomalías como las apuntadas, los responsables académicos tienen al menos tres grandes planos diferenciados, aunque no del todo independientes, de evaluación e intervención. Estos planos son:

- El diseño de asignaturas del plan de estudios
- Los contenidos de las asignaturas
- Los niveles de dificultad de los contenidos tratados

En el primer plano, en el de las *asignaturas del plan de estudios*, existen amplias posibilidades de intervención de los responsables institucionales, a través de decisiones de los diferentes órganos colegiados, pudiéndose introducir racionalidad académica y profesional, criterios a los que se hace referencia de manera constante. Sin embargo, a lo largo de este proceso suelen explotar innumerables contradicciones y tensiones internas, frecuentemente no debidas a razones académicas y pedagógicas, sino a intereses particulares, generalmente relacionados con la creación de necesidades de dotación de nuevo profesorado. Sin ningún género de dudas, la coherencia y racionalidad académica y formativa de los planes de estudio universitarios debe ser uno de los objetos prioritarios de la evaluación institucional y algo sobre lo que las comisiones docentes, las juntas de departamento, de centro y de gobierno no pueden pasar por encima escudándose en el objetivo de evitar conflictos y dificultades.

En un informe solicitado al Vicerrector de mi Universidad responsable de la elaboración y reforma de los planes de estudio (López, 2000), se ofrece desde dicha perspectiva, una revisión crítica del conjunto de elementos que dificultan o pueden dificultar el proceso aludido y que, consecuentemente, requieren atención evaluadora. La clave de los desajustes se ubica entre la concepción flexible de las directrices generales y el contexto universitario real en el que tienen que desarrollarse y entroncarse los planes.

Entre los elementos de desequilibrio señalados por el informe se pueden citar: a) la diversidad de criterios de aplicación según titulaciones, b) las disfunciones en cuanto a peso de troncalidad de las diferentes materias, fruto del diferente potencial académico de las áreas, más que de la racionalidad formativa, c) las nuevas dificultades para la gestión académica, disgregación de grupos de estudiantes, configuración de la optatividad y la libre elección, vinculación de áreas de conocimiento, etc., d) compresión

403

excesiva de contenidos al aplicar el modelo cuatrimestral, e) uso arbitrario de los prerequisitos de acceso a determinadas asignaturas, f) dificultades organizativas del *practicum*, con creciente necesidad de profesionales-tutores que demandan compensación por su colaboración, g) aumento de la complejidad organizativa de los segundos ciclos debido a la transversalidad y a la diversidad de opciones de origen de los estudiantes, h) falta de criterios comunes sobre la duración de las titulaciones e i) carencia y dificultad de establecer mecanismos precisos de adaptación curricular y de convalidadión de créditos.

Problemas como los apuntados, que vemos son motivo de gran preocupación en el momento presente para nuestro responsables académicos, deben ser obligatoriamente analizados en los programas de evaluación curricular, con la correspondiente oferta de posibles soluciones alternativas.

En cuanto al segundo plano, el relativo al *contenido de las asignaturas*, las responsabilidades son principalmente departamentales e interdepartamentales, siendo los problemas más de sensatez académica y pedagógica, que de defensa laboral o de influencia. El gran reto es el de convencer a los profesores que en los programas se debe introducir de manera preferente lo más formativo y lo más útil, y no introducir nuevos contenidos sin desarrollar previamente los anteriores de manera adecuada y sin lograr los objetivos previstos.

Ningún evaluador puede olvidarse de que la linealidad en el tratamiento de contenidos en muchas de nuestra asignaturas es un problema grave; se selecciona poco y se actúa como si toda la materia fuera igual de importante y formativa. Los departamentos y los grupos docentes deben hacer autocrítica en tal sentido, aligerando los programas de contenidos menos importantes. En este terreno, los profesores noveles deben contar con el asesoramiento departamental. También debe resaltar la evaluación la dimensión de utilidad de los programas de las asignaturas, pero no sólo reforzando el contenido práctico del diseño curricular, sino enfatizando la orientación de utilidad en la selección y el tratamiento de los contenidos. Lo mismo podríamos decir de la necesidad de coordinación intra e interdepartamental a la hora de distribuir los contenidos entre asignaturas y mejorar la eficiencia docente. Se trata, por ejemplo, de cuidar las jerarquias conceptuales en el tratamiento temporal de los contenidos, de evitar que un mismo tema se trate innecesariamente varias veces por distintos profesores y, también, de asegurar que no se deje de tratar algún tema importante por problemas de coordinación.

El tercer plano de evaluación en el ámbito curricular es el relativo a los *niveles de tratamiento y dificultad* de los contenidos de cada asignatura que, en buena medida, quedan ceñidos al juicio de cada profesor y, a lo sumo, a la coordinación del grupo de profesores de la materia. En este plano resulta muy difícil entrar como acción institucional, lo que es un problema muy grave, sobre todo si se confirma la hipótesis defendida por muchos evaluadores, de que este aspecto es uno de los más determinantes en el denominado fracaso escolar.

La reflexión final en este importante ámbito, el curricular, es que la diversidad entre centros, departamentos, titulaciones, profesores, asignaturas,...es enorme y que suele ser no sólo erróneo, sino muy peligroso, generalizar los hallazgos de evaluaciones específicas. Hay que tratar los problemas de manera diferenciada según la situación.

3.4. La docencia, sus recursos y sus métodos

El último eslabón de esta cadena de ámbitos de evaluación y mejora de nuestra enseñanza universitaria es el de la docencia en sentido más estricto, esto es, el de la interacción didáctica entre alumnos y profesores. La docencia está muy condicionada por los aspectos curriculares discutidos anteriormente, pero siempre queda un margen de maniobra estrictamente didáctico que tiene que ver sobre todo con los enfoques metodológicos, con las técnicas didácticas y con los recursos de apoyo empleados en las aulas. En definitiva, estamos hablando de la actuación del profesor y del contexto pedagógico en el que se mueve.

Desde la perspectiva evaluadora podemos hablar, al menos, de dos planos distintos aunque interconexos, la *actuación didáctica* de los profesores y los *medios y condiciones* que facilitan o dificultan ciertas conductas pedagógicas. Reflexionemos sobre ambos planos.

La experiencia en tareas de formación del profesorado nos dice que el cambio de conducta didáctica de los profesores no es, precisamente, empresa fácil, pero también nos dice que es un arma poderosa para la mejora de la enseñanza. El problema no está tanto en los cambios en la preparación didáctica del profesorado, pues son relativamente sencillos, sino en la modificación de su capacidad, de su concepción, de su actitud y de su conducta pedagógica, algo muchísimo más costoso.

No es mi intención, ni tampoco tengo espacio, entrar en una reflexión profunda sobre la evaluación y mejora del profesorado universitario, pero nos conviene recordar que por lo dicho anteriormente, por su dificultad y por su importancia, estamos hablando de un elemento crítico pues, en definitiva, vemos a los sistemas de evaluación y compensación de los profesores como palancas de cambio (Malen, 1999). Tampoco podemos soslayar los problemas que surgen ante la gestión de incentivos y penalizaciones personales como una de las potenciales consecuencias de la evaluación (Escudero, 2000a).

Sabemos que las penalizaciones personales inciden en el nivel más elemental de la calidad docente, pudiendo evitar incumplimientos laborales y corregir defectos de funcionamiento básico, pero no son un instrumento de generación de calidad en sentido proactivo. Sobre los incentivos personales hemos de convenir que su potencial de mejora teórico no siempre se hace real, pues su articulación práctica suele ser conflictiva. Uno de los principales dilemas está en saber si el incentivo verdaderamente estimula al que lo consigue o, más bien, desestimula todavía más al que no lo consigue.

Las encuestas de evaluación de la docencia y de los profesores que utilizamos en las universidades se encuentran inmersas en esta situación conflictiva (Escudero, 1999a), si se entienden como un indicador objetivo para la penalización más que como un instrumento de diagnóstico con el que promover cambios y mejoras docentes. Para esta segunda tarea las encuestas son un elemento imprescindible, pero para la primera su utilidad es muy limitada y, en todo caso, tendrían valor complementario de otras necesarias fuentes de información. La evaluación institucional debe atender este importante problema y sugerir reformas de estos procesos, pues el estado actual de

cosas, además de ser altamente improductivo, es un foco de desprestigio de la propia evaluación como vehículo para la mejora de la enseñanza.

Los *recursos docentes* tienen, sin duda, un tratamiento evaluativo mucho menos complejo que el del profesorado y bastante sencillo en cuanto a intervenciones posteriores. Aquí debemos apoyarnos en los clásicos esquemas de la *evaluación de necesidades*, pero marcando siempre prioridades entre las detectadas, según niveles y tipos (Witkin y Altschud, 1995). La eficiencia es un criterio imprescindible a cuidar por el evaluador en todo plan de recursos o equipamientos docentes, pues no es verdadera necesidad todo lo que no se tiene o es deficiente, además debe existir la perspectiva de su utilización y esta detección no siempre es sencilla.

En un momento de efervescencia de la utilización de tecnologías de la información y la comunicación en las aulas universitarias, con mucha innovación metodológica en marcha y también con cierta confusión en muchos profesores entre calidad docente y nuevas tecnologías, los expertos en evaluación educativa tienen la obligación profesional de introducir elementos de racionalidad en esta especie de carrera desenfrenada, alertando de la necesidad de que se cumplan algunas condiciones que van más allá de los simples cambios técnico-metodológicos.

No tengo ninguna duda de que la experimentación y la innovación didácticas son importantes caminos para la mejora de la enseñanza, pero no la aseguran por sí solas. Deben ir acompañadas de efectos positivos en los resultados académicos, en las actitudes y conductas de los estudiantes, etc. Por ello, *la innovación debe ser evaluada*, con planes de seguimiento que analicen críticamente sus efectos. Desgraciadamente, esta evaluación es escasa y tengo la impresión fundada de que muchas de las innovaciones que se están introduciendo son didacticamente muy discutibles o, incluso, contraproducentes. El enorme potencial que tienen las denominadas nuevas tecnologías, debe ser aprovechado de manera crítica, asegurándo siempre su eficiencia pedagógica. Por ello es imprecindible la evaluación.

4. UN APUNTE, PARA FINALIZAR

Estas páginas quieren ofrecer un conjunto de reflexiones sobre los ámbitos de evaluación y mejora de la enseñanza universitaria actual que, a mi entender, son de obligada preocupación, si se toma la perspectiva del responsable académico e institucional.

Es una obviedad señalar que nuestra enseñanza universitaria está inmersa en un contexto que es sugerente por la aparición de nuevos enfoques y recursos metodológicos, pero que tiene muchos elementos de ineficacia. Los retos y los problemas planteados son muchos y variados y las respuestas y soluciones no vienen fácilmente a la mano.

Apoyado en la experiencia y en la literatura profesional se me fortalece cada vez más la idea de que son muchos y muy variados los elementos que influyen sobre la enseñanza en general y sobre la universitaria en particular y que, en conjunto, ninguno de ellos tiene un carácter determinante; son muchas las aportaciones, frecuentemente modestas, las que construyen una enseñanza universitaria de calidad.

La guinda de la cuestión es el amplio abanico de situaciones y problemas, con una etiología tan diversa entre centros, titulaciones, departamentos, cursos, asignaturas,

profesores, estudiantes, etc., que, aunque existieran, hacen inútiles los tratamientos generales. Además, los puntos de responsabilidad y de autonomía son tantos, que resulta impensable una aplicación en bloque.

Personalmente creo que la mejor opción para los responsables universitarios y para los profesionales de la educación es la de procurar extender cada vez más, dentro de la institución y en todas y cada una de sus unidades, la cultura de la evaluación, de la autocrítica y la mejora continua, con intervenciones que incidan en terrenos como los aquí analizados. El estímulo y los medios aportados marcarán el avance.

BIBLIOGRAFÍA

De Miguel, M. (1999). La evaluación de la enseñanza. Propuesta de indicadores para las titulaciones, en *Indicadores en la Universidad: información y decisiones*, Consejo de Universidades, MEC, Madrid, 413-430.

Escudero, T. (1999). Indicadores del rendimiento académico: Una experiencia en la Universidad de Zaragoza, en *Indicadores en la Universidad: información y decisiones*, Consejo de Universidades, MEC, Madrid, 251-262.

Escudero, T. (1999a). Los estudiantes como evaluadores de la docencia y de los profesores: nuestra experiencia, *Revista Interuniversitaria de Formación del Profesorado*, 34, 69-86.

Escudero, T. (2000). Evaluación de centros e instituciones educativas: Las perspectivas del evaluador, en González, D. y otros (Coords.), *Innovación en la escuela y mejora de la calidad educativa*, Grupo Editorial Universitario, Granada, 57-76.

Escudero, T. (2000a). Evaluación y motivación del profesorado universitario, *I Congreso Internacional «Docencia Universitaria e Innovación»*, 26-28 de junio, Barcelona.

Escudero, T. y otros, (1997). *Nuestra docencia según los estudiantes*, Vicerrectorado de Evaluación y Mejora de la Enseñanza, Universidad de Zaragoza.

Escudero, T. y otros, (1999). *Resultados académicos por titulaciones. Algunos indicadores de rendimiento*, Vicerrectorado de Evaluación y Mejora de la Enseñanza, Universidad de Zaragoza.

Escudero, T. y otros, (2000). *Evaluación de las prácticas en la licenciatura de Medicina*, Informes 48, ICE-Universidad de Zaragoza.

Ferrández, R. y otros, (1999). Problemas en la interpretación del rendimiento de los estudiantes como indicador de calidad de una institución, en *Indicadores en la Universidad: información y decisiones*, Consejo de Universidades, MEC, Madrid, 263-274.

Fishman, D.B. (1989). Pragmatic behaviorism: Saving and nurturing the baby, en Fishman D.B. y otros (Eds.), *Paradigms in behavior therapy: Present and promise*, Springer, New York, 254-293.

Fishman, D.B. (1991). An introduction to the experimental versus the pragmatic paradigm in evaluation, *Evaluation and Program Planning*, 14, 4, 353-363.

Gredler, M.E. (1996). *Program Evaluation*, Merril/Prentice-Hall, Englewood Cliffs, N.J.

López, M. (2000). *Informe personal*, Vicerrectorado de Ordenación Académica, Universidad de Zaragoza.

Malen, B. (1999). On rewards, punishments, and possibilities: Teacher compensation as an instrument for education reform, *Journal of Personnel Evaluation in Education*, 12, 4, 387-394.

Municio, P. (1992). La evaluación segmentada de los programas, *Bordon*, 43, 4, 375-395.

Pérez Juste, R. (1995). Evaluación de programas educativos, en Medina, A., y Villar, L.M. (Coord.), *Evaluación de programas educativos, centros y profesores*, Editorial Universitas, S.A., Madrid, 73-106.

Segers, M. y Dochy, F. (1996). Quality assurance in higher education: Theoretical considerations and empirical evidence, *Studies in Educational Evaluation*, 22, 2, 115-137.

Witkin, B.R. y Altschuld, J.W. (1995). Planning and conducting needs assessments. A practical guide, Sage, Thousand Oaks, Ca.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 417-432

LA EVALUACIÓN DE LA ACTIVIDAD DOCENTE EN LA UNIVERSIDAD: ENTRE EL SUEÑO Y LA REALIDAD

Gregorio Rodríguez Gómez
Universidad de Cádiz

RESUMEN

En este trabajo se ofrecen algunas evidencias de las aportaciones que, desde una perspectiva cualitativa, se pueden ir ofreciendo a la difícil tarea que supone la evaluación de la docencia en el contexto universitario. Para ello se recurre a una breve presentación de la experiencia llevada a cabo durante el curso 97/98 en la Universidad de Cádiz, con el objetivo de profundizar en las posibilidades y limitaciones de las técnicas cualitativas para abordar la evaluación de la actividad docente universitaria.

SUMMARY

In this paper are offered some evidences about contributions that a qualitative perspective can offer to the diffiult task of teaching evaluation at university context. This paper make a short presentation of the experience carried out during 1997/1998 at University of Cadiz with the objective to deep on possibilities and limitations of qualitative tools, in order to evaluate the university teaching activities.

INTRODUCCIÓN

Plantear la evaluación de la actividad docente universitaria significa adentrarse en un contexto tremendamente complejo y arduo que requiere ser afrontado desde una gran multiplicidad de perspectivas. En ocasiones, con más frecuencia de la debida, se utilizan los términos evaluación de la enseñanza o de la docencia y evaluación del profesorado como sinónimos. Un ejemplo de ello lo constituye el trabajo de Mateo (2000).

Este hecho puede tener cierta justificación en el contexto de niveles educativos no universitarios, donde la actividad docente cubre la mayor parte de la actividad profesional del profesorado. Pero en un contexto universitario las tres funciones básicas asignadas al profesorado (docencia, investigación y gestión y extensión universitaria), hace necesaria que éstas se diferencien claramente, máxime si nos situamos desde una perspectiva evaluativa, pues previamente ha de determinarse de forma clara y precisa el objeto de evaluación.

En el presente trabajo vamos a centrar nuestra atención en la evaluación de la actividad docente del profesorado universitario, hecho éste que ha sido objeto de tratamiento en numerosos trabajos, entre los que quisiéramos destacar en este momento las aportaciones de Aylet y Gregory (1996), De Miguel (1989, 1991, 1996, 1997 y 1998), Dunkin (1995), Zabalza (1999) o Zúñiga (1997), entre otras. Más concretamente, ofreceremos algunas evidencias de las aportaciones que, desde una perspectiva cualitativa, se pueden ir ofreciendo a la difícil tarea que supone la evaluación de la docencia en el contexto de nuestras universidades. Para ello recurriremos a una breve presentación de la experiencia llevada a cabo durante el curso 97/98 en la Universidad de Cádiz, con el objetivo de profundizar en las posibilidades y limitaciones de las técnicas cualitativas para abordar la evaluación docente.

LA EVALUACIÓN DE LA ACTIVIDAD DOCENTE UNIVERSITARIA

Sobre la base del Real Decreto 1086/1989 de 28 de agosto, y la Resolución de 20 de junio de 1990 (BOE del 30 de junio), en las universidades españolas se comenzó a desarrollar un modelo de evaluación que «no ha tenido ningún tipo de consecuencias a efectos de la continuidad y promoción del profesorado» (De Miguel, 1998: 68), lo que ha venido a poner en duda el modelo y, como consecuencia, las técnicas e instrumentos que se vienen utilizando.

Desde una perspectiva más comprensiva de la realidad, De Miguel (1998) propone un modelo de evaluación que tiene dos características fundamentales: centrado en el profesor y basado en la recogida de evidencias múltiples.

Desde la consideración del profesorado universitario, la evaluación de su actividad docente se viene considerando como algo externo sin implicación del propio profesor, trasladando la responsabilidad de su realización a agentes externos. A esta concepción de la evaluación ha colaborado en gran medida los procedimientos que se han puesto en práctica, los cuales se han basado casi exclusivamente en la utilización de cuestionarios o escalas que son cumplimentadas por parte del alumnado. Así pues, la evaluación de una actividad tan compleja como la docencia se realiza tomando en cuenta una única fuente de información (el alumnado) y una única fuente de evidencias (el cuestionario).

Si partimos de la base de considerar la evaluación de la docencia universitaria como un medio para la mejora, es el propio profesor el que ha de preocuparse por asumir la responsabilidad de recoger la información adecuada para reflexionar sobre su propia actividad, así como para dar cuenta de la misma.

En torno a la información a recabar, el modelo propuesto por De Miguel (ver cuadro 1) considera cuatro dimensiones básicas, sobre cada una de las cuales el profesora-

CUADRO 1
MODELO DE ACTUACIÓN DOCENTE EN EL CONTEXTO INSTITUCIONAL (DE MIGUEL, 1998: 79-80)

1. CONTEXTO INSTITUCIONAL	
Características y preocupación por la docencia en el centro y/o titulación	— tamaño, ratios profesor/alumno/grupos, etc. — calificaciones de entrada, servicios de ayuda al estudiante, etc. — preocupación por la calidad de la enseñanza, estímulos para la docencia, etc.
2. CONDICIONANTES INICIALES DE LA ACTIVIDAD DOCENTE	
Volumen y rango de la enseñanza impartida (carga docente)	— cumplimiento de la gestión ordinaria — cantidad del teimpo dedicado a la enseñanza — número y complejidad de las materias impartidas — número y diferencias en el alumnado
Actividades de formación y perfeccionamiento docente	— asistencia a cursos, congresos y eventos sobre formación pedagógica — participación en grupos de trabajo y asociaciones sobre docencia — publicaciones sobre aspectos relacionados con la enseñanza
Preparación de las clases	— claridad y adecuación de las metas y objetivos de cada asignatura — actualización de los contenidos — preparación de materiales y equipos
Coordinación de la enseñanza en el ámbito institucional	— actividades de coordinación de la docencia — responsabilidades institucionales — participación en programas orientados a la mejora de la actividad docente
3. DESARROLLO Y DEDICACIÓN A LOS PROCESOS DE ENSEÑANZA-APRENDIZAJE	
Valoración de la competencia docente	— evidencias sobre la eficacia en las clases — evidencias sobre la adecuación de los métodos a los objetivos educativos (teóricos, prácticos, etc.)
Innovación en la enseñanza	— innovaciones en los contenidos curriculares — innovaciones metodológicas — interdisciplinariedad y trabajo en equipo — innovaciones relativas a programas nacionales e internacionales
Relaciones con los estudiantes y servicios de apoyo	— disponibilidad fuera de las clases — sistemas de orientación y consulta
Procedimientos de evaluación y autoevaluación	— evidencias sobre los métodos utilizados — innovación en las técnicas de evaluación — sistemas de feedback utilizados — reflexión sistemática y regular sobre la práctica docente — uso de la evaluación de los alumnos y colegas — revisión de la enseñanza en relación con: a) los fines y objetivos de la titulación b) los contenidos curriculares de la materia
4. INDICADORES RELATIVOS A LOS RESULTADOS E IMPACTOS	
Resultados académicos de los alumnos	— tasas de éxito, fracaso y abandono
Reconocimiento dentro de la institución	— valoraciones de los estudiantes y colegas — evaluaciones de los responsables académicos — cargos docentes desempeñados por elección
Reconocimientos fuera de la institución	— invitaciones para impartir docencia en otros centros: a) conferencias, seminarios, cursos nacionales e internacionales b) estancias regulares en otros departamentos universitarios (máster, doctorados, posgrados, etc.) — servicios profesionales a otras universidades y organizaciones en tanto como evaluador — becas y contratos conseguidos para mejorar la enseñanza

BEST COPY AVAILABLE

do habrá de presentar las evidencias pertinentes. Si analizamos cada una de estas dimensiones podremos ver claramente que en modo alguno podrán aportarse evidencias utilizando una única fuente de información o un simple instrumento de recogida de información.

Desde una consideración estrictamente metodológica para recoger la información necesaria, sobre la que basar las evidencias en cada una de las dimensiones planteadas en el modelo de De Miguel (1998), se podría utilizar una multiplicidad de métodos y técnicas. Concretamente, si centramos la atención a nivel de aula, en el cuadro 2 podemos observar una primera clasificación de éstos.

Situándonos en esta perspectiva, en la que se tienen en cuenta diversos momentos, métodos y técnicas de recogida de información, que ayudarán a comprender una reali-

CUADRO 2
MÉTODOS Y TÉCNICAS PARA LA EVALUACIÓN DE LA ACTIVIDAD DOCENTE

Métodos	**Técnicas**
Autoevaluación	— Feedback del propio comportamiento docente — Formularios de autovaloración — Materiales de autovaloración — Autoinformes — Observación en clase de colegas — Revisión de compañeros
Evaluación por los compañeros	— Con observación — Sin observación
Evaluación por supervisores	— Observación en el aula
Evaluación por los alumnos	— Aprendizaje de los alumnos — Cuestionarios de evaluación — Entrevistas grupales
Otros	— Valoraciones de administradores — Portafolio

dad compleja y facilitarán la toma de decisiones, con el fin de mejorar la actividad docente, desde la Unidad para la Calidad de la Universidad de Cádiz se están realizando ciertas aproximaciones para evaluar la actividad docente del profesorado, una de las cuales comentamos seguidamente.

UNA EXPERIENCIA DE EVALUACIÓN DE LA DOCENCIA DESDE UNA PERSPECTIVA CUALITATIVA

Durante el curso 1996/1997 la evaluación de la docencia en la Universidad de Cádiz se había realizado en convocatoria pública de carácter voluntario. Así, cualquier profesor interesado solicitaba su participación y desde la Unidad para la Calidad se arbitraban los medios necesarios para recabar la opinión de los alumnos sobre la actividad docente del profesorado, de tal suerte que los cuestionarios de opinión eran cumplimentados por parte de los alumnos bajo las orientaciones de encuestadores

independientes y, si así lo deseaba, en presencia del propio profesor, guardando en todo momento la confidencialidad necesaria en estos casos.

Este procedimiento permitió una mayor fiabilidad y credibilidad de los resultados obtenidos, pero seguía sin aportar informaciones que permitieran comprender mejor la actividad concreta de cada profesor.

En un intento de superación de este modelo, durante el curso 1997/1998 se inició desde la Unidad para la Calidad de la Universidad de Cádiz una experiencia de evaluación de la docencia, con el fin de experimentar un procedimiento de evaluación que sirviera para introducir nuevos métodos de evaluación para una mejor aproximación a la realidad docente de cada profesor.

Así, se realizó una nueva convocatoria pública entre todo el profesorado que durante el curso 1996/1997 había participado en el proceso de evaluación de la docencia siguiendo el modelo de encuesta, ofreciéndole participar en una experiencia que implicaría un mayor seguimiento de su actividad docente. La respuesta fue muy positiva y se tuvo que proceder a una selección aleatoria de participantes, pues con los medios humanos que se contaban en ese momento no se podía dar respuesta a todas las peticiones. En total se seleccionaron 20 profesores de diferentes facultades.

En la figura 1 podemos observar el carácter complementario y global que se pretende con este nuevo modelo de evaluación de la docencia. En definitiva lo que se perseguía era recabar información sobre la actividad docente del profesorado implicando al mismo en el proceso de una forma más directa y tomando otras fuentes de información complementarias. Así la opinión del profesorado y de sus alumnos se recogería a partir de entrevistas, informaciones que se complementarían con las observaciones realizadas por parte de los observadores externos y la confrontación con otros documentos tales como programas de las asignaturas, lecturas, etc.

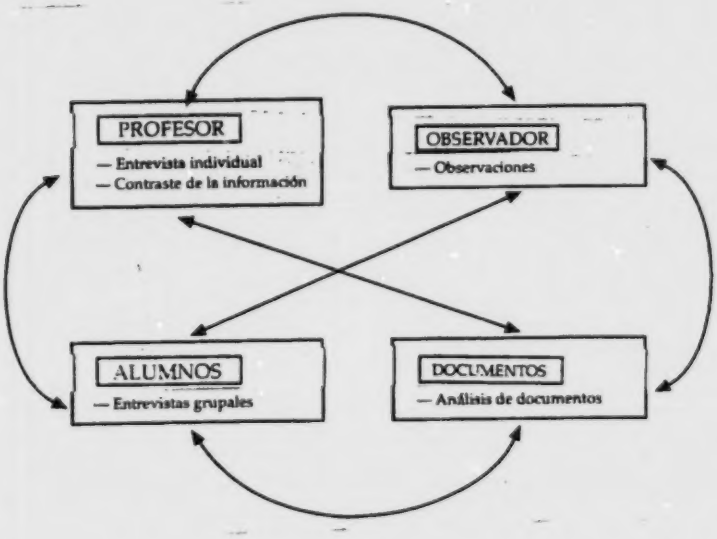


Figura 1
Fuentes de información para la evaluación de la docencia

El proceso que se siguió se sintetiza en la figura 2, en la que podemos observar las distintas fases que se desarrollaron en la ejecución del mismo.

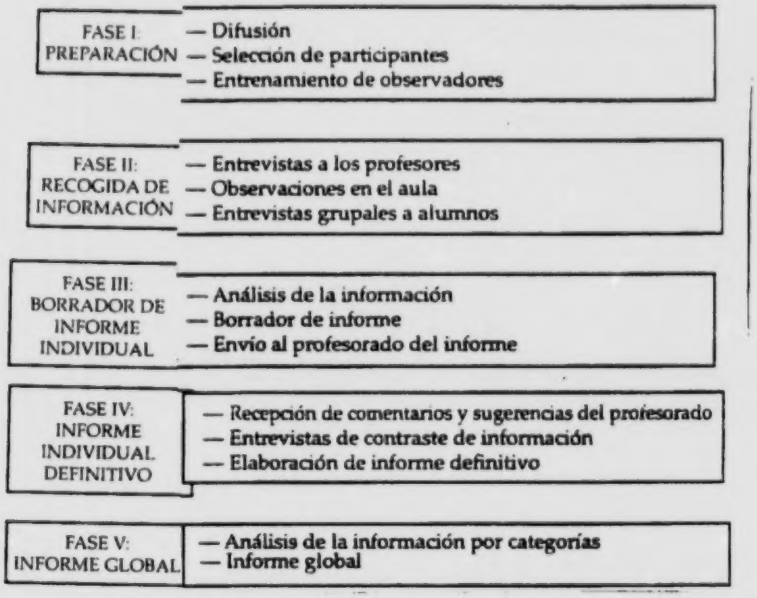


Figura 2
Fases en el proceso de evaluación de la docencia

Primera fase: *Preparación*

Difusión del proyecto:

A lo largo del primer trimestre del curso 1997/1998 se difundió una carta de presentación del proyecto entre el profesorado de la Universidad de Cádiz que durante el curso anterior había realizado la evaluación de su actividad docente a través del procedimiento de encuesta al alumnado.

Selección del profesorado participante:

Una vez recibidas las solicitudes de participación en el proyecto se procedió a seleccionar un total de 20 profesores, a los que se les comunicó que a partir del mes de marzo de 1998 que se procedería a realizarle entrevistas, observaciones de sus clases y entrevistas a sus alumnos.

Entrenamiento de los observadores:

Los observadores que participaron en esta experiencia eran todos alumnos de la Facultad de Ciencias de la Educación que cursaban su último año de la Licenciatura de Psicopedagogía, y concretamente cursaban la asignatura de Metodología de la Investigación Cualitativa.

414

Antes de iniciar el proceso de evaluación se mantuvieron sesiones de formación para familiarizarse con las técnicas cualitativas que iban a ser utilizadas posteriormente. Al finalizar esta etapa formativa se constituyeron un total de 20 equipos de observadores formados por dos o tres alumnos.

Segunda Fase: *Recogida de Información*

El proceso de recogida de información se llevó tras acordar el mismo entre el profesor implicado y el equipo de observación, tras lo cual se realizaron las siguientes actividades:

Realización de entrevistas a los profesores

Las entrevistas se realizaron a los profesores bajo el formato de entrevista semiestructurada, de acuerdo con el guión de entrevista que se presenta en el Anexo I. Estas entrevistas fueron grabadas en audio y transcritas posteriormente para facilitar su análisis posterior.

A lo largo de la entrevista y de la observación en el aula se le requirió al profesorado toda la documentación que considerara de interés sobre la asignatura (programa, apuntes, evaluaciones ...etc.)

Realización de observación en las aulas

Durante dos o tres sesiones se realizaron las observaciones en las aulas de los profesores, tomando como guía de observación la que se presenta en el Anexo II. El registro de estas observaciones se llevó a cabo a través de notas de campo, que eran posteriormente archivadas en formato electrónico.

Realización de entrevistas grupales

La opinión del alumnado se recabó a través de la realización de entrevistas grupales, de acuerdo con la guía de entrevista que se adjunta en el Anexo III. Los alumnos entrevistados fueron seleccionados bajo el criterio de representatividad de los mismos y participación e implicación en las actividades del aula, de forma que pudieran aportar información de calidad.

Tercera fase: *Elaboración de borrador de informe individual de evaluación*

En esta tercera fase el objetivo fundamental era realizar un informe individualizado de la actividad docente de cada uno de los profesores, sobre la base de toda la información recabada en la fase anterior. Para ello se siguieron las siguientes etapas:

Análisis de la información

En primer lugar cada equipo de observación realizó un análisis de la información recabada en la etapa anterior, siguiendo en este sentido el proceso general de análisis (Rodríguez y otros, 1996) y utilizando como sistema de categorías el presentado en el Anexo IV.

Realización del borrador de informe

Sobre la base de las entrevistas, las observaciones y el análisis de los documentos cada equipo de observación realizó un primer informe sobre la actividad docente, que con una longitud que osciló entre 7 y 12 páginas seguía básicamente la siguiente estructura:

— Introducción
— Contexto (Aula, plan de estudios, asignatura)
— Metodología (planificación, objetivos, metodología en el aula)
— Evaluación (técnicas, instrumentos, criterios ...)
— Relaciones con los alumnos
— Valoración global
— Propuestas de mejora

Envío al profesorado del borrador de informe

Una vez elaborado el borrador de informe éste se enviaba al profesorado para que procediera a su lectura y análisis y, posteriormente en la siguiente fase, proceder a su contraste con los observadores.

Cuarta fase: *Elaboración del informe individual de evaluación definitivo*

El objetivo de esta cuarta fase era el de elaborar el informe final de evaluación sobre la actividad docente de cada uno de los profesores, sobre la base del consenso a partir de las informaciones y documentos analizados en las fases anteriores. Las etapas que se sucedieron fueron:

Recepción de comentarios y sugerencias al borrador

Una vez recepcionado por parte del profesor el informe de evaluación, y tras la lectura del mismo, si tenía que realizar alguna aclaración, aportación o crítica éstas eran enviadas al equipo de observación.

Contraste de la información

Para contrastar las sugerencias aportadas por cada uno de los profesores, cada equipo de observación mantuvo una última entrevista con cada profesor, con la finalidad de aclarar las dudas o recomendaciones realizadas y llegar a un consenso sobre el contenido del informe final de evaluación.

Elaboración del informe definitivo

Una vez alcanzado el consenso, por parte del equipo de observación se realizó el informe final de la actividad docente de cada profesor, enviando el mismo a cada profesor participante.

Quinta fase: *Elaboración de informe global*

En esta última fase del proceso de evaluación la unidad de análisis y evaluación cambiaba el foco de atención, pasando de la atención prestada hasta ese momento a

la actividad individual de cada profesor, a las categorías de análisis que se consideraron más importantes: contexto de la docencia, metodología, evaluación y relaciones con los alumnos. El objetivo era disponer de un informe que aportara una visión global de la actividad docente del profesorado, para lo cual se siguieron las siguientes etapas:

Análisis de la información sobre la base del sistema de categorías

En este momento se llevó a cabo un análisis de todo la información recopilada (entrevistas, notas de campo y documentos) tomando como foco de atención las principales categorías. Las cuestiones que guiaron este proceso de análisis eran:

— ¿Cuáles son las características del contexto en el que los profesores desempeñan su actividad docente?
— ¿Cuáles son las principales características de la metodología seguida por parte del profesorado en sus aulas?
— ¿Qué caracteriza al proceso de evaluación que sigue el profesorado?
— ¿Cómo podrían caracterizarse las relaciones de los profesores con el alumnado?

Elaboración de informe final global por categorías

Tomando como base cada una de las principales categorías analizadas se realizó un informe global de la actividad docente del profesorado implicado en esta experiencia de evaluación.

A MODO DE CONCLUSIÓN

Como han puesto de manifiesto diferentes autores, el modelo de evaluación de la actividad docente puesto en práctica hasta el momento en el contexto de nuestras universidades españolas es a todas luces insuficiente para alcanzar una mínima comprensión de esa realidad tan compleja.

La experiencia expuesta en estas páginas realizada en la Universidad de Cádiz y otras similares, como las llevadas a cabo por parte de Álvarez Rojo (1999) en la Universidad de Sevilla, ponen de manifiesto que la puesta en práctica de modelos alternativos como el planteado por De Miguel (1999) son factibles.

La cultura de la evaluación y la preocupación por una mejora de la actividad docente viene generalizándose cada vez más, como lo ponen de manifiesto las numerosas jornadas que desde diferentes áreas de conocimiento se dedican de forma monográfica a la presentación de aportaciones en el terreno de la docencia universitaria.

Así pues, en estos momentos disponemos de las técnicas e instrumentos, así como de la cultura evaluativa suficiente como para poder abordar eficazmente estas propuestas alternativas de evaluación de la actividad docente, de tal forma que incidan en mejoras significativas de la calidad de la misma. La cuestión radica en que desde una perspectiva política se tomen las decisiones oportunas que faciliten los medios y recursos necesarios para ello.

BIBLIOGRAFÍA

Álvarez Rojo, V. (1999). *Profundizando en la calidad de la enseñanza: aportaciones de los profesores mejor evaluados de la Universidad de Sevilla.* Sevilla: Servicio de Publicaciones - Vicerrectorado de Calidad - ICE.

Aylet, R. y Gregory, K. (1991). *Assessment for Excellence.* Nueva York: MacMillan.

De Miguel, M. (1989). *Criterios para la evaluación del profesorado universitario.* Oviedo: KRK Ediciones.

De Miguel, M. (1991). Utilización de indicadores en la evaluación de la docencia universitaria. En M, De Miguel, J. Mora y S. Rodríguez (Eds.), *La evaluación de las instituciones universitaria.* Madrid: Consejo de Universidades / MEC.

De Miguel, M. (1996). La evaluación de la función docente del profesorado universitario. En J.M. Rodríguez (Ed.), *Actas del Seminario de Formación y Evaluación del Profesorado Universitario.* Huelva: Servicio de Publicaciones de la Universidad de Huelva / ICE.

De Miguel, M. (1997). Evaluación y reforma pedagógica en la enseñanza universitaria. En P. Apodaca y C. Lobato (Eds.), *Calidad en la Universidad: Orientación y evaluación.* Barcelona: Laertes.

De Miguel, M. (1998). La evaluación del profesorado universitario. Criterios y propuestas para mejorar la función docente. *Revista de Educación,* (315), 67-83.

Dunkin, M. (1995). Concepts of teaching and teaching excellence in Higher Education. *Higher Education Research and Development, 14* (1), 21-33.

Ellis, R. (Ed.) (1993). *Quality Assurance for University Teaching.* Londres: The Society for Research into Higher Education / Open University Press.

Elton, L. (1996). Criteria for Teaching Competence and Teaching Excellence in Higher Education. En R. Aylet y K. Gregory (Eds.), *Assessment for Excellence.* Nueva York: MacMillan.

Mateo, J. (2000). La evaluación del profesorado y la gestión de la calidad de la educación. Hacia un modelo comprensivo de evaluación sistemática de la docencia. *Revista de Investigación Educativa, 18* (1), 7-36.

Zúñiga, R. (1997). La evaluación en la acción docente. En P. Apodaca y C. Lobato (Eds.), *Calidad en la Universidad: Orientación y evaluación.* Barcelona: Laertes.

ANEXO I:
GUIÓN DE ENTREVISTA SEMIESTRUCTURADA AL PROFESORADO

1. **Biografía personal**
   - ¿Puede contarnos a grandes rasgos su trayectoria académica y profesional?
   - ¿Qué le movió a elegir esta carrera?
2. **Actividad profesional**
   - ¿Qué asignaturas, grupos y alumnos tiene a su cargo?
   - ¿Podría describirnos una semana típica de trabajo?
   - ¿Está o ha estado realizando alguna investigación?
   - ¿Influye en su propia docencia aquello sobre lo que investiga?
   - ¿Cómo coordina su labor docente/ investigadora y administrativa?
3. **Planificación de la enseñanza**
   - ¿Cómo planifica la docencia de su asignatura?
   - ¿Cuáles son sus criterios de cara a seleccionar objetivos y contenidos y su distribución en el tiempo?
   - ¿Existe coordinación docente con otros compañeros?
   - ¿Qué aspectos de la docencia se discuten o negocian con los alumnos?
   - Existen diferencias entre la planificación de una clase teórica y una práctica?
4. **Organización de la clase**
   - ¿Cómo se estructura el desarrollo de una clase?
   - ¿Organiza a los alumnos en la clase de algún modo particular?
   - ¿Utiliza algún tipo de recurso para impartir una clase (pizarra, retro, vídeo ... etc.)?
   - ¿Qué tipo de cosas planea de cara a motivar y conseguir el interés de los alumnos?
5. **Método de enseñanza**
   - ¿Cómo describiría el método de enseñanza que sigue en su clase?
6. **Evaluación**
   - ¿Cuáles son sus criterios de evaluación?
   - ¿Qué evalúa?
   - ¿Cómo evalúa el aprendizaje de sus alumnos?
   - ¿Influyen los resultados de la evaluación para modificar su práctica docente?
7. **Tutoría**
   - ¿Cómo enfoca las sesiones de tutoría con sus alumnos?
   - ¿Asisten los alumnos a tutorías?, ¿para qué?
8. **Relaciones profesionales**
   - ¿Qué comentarios piensa que podría hacerme un alumno suyo sobre usted? ¿Y sus compañeros?
   - ¿Cómo describiría su relación con sus alumnos? ¿Y con sus compañeros?
   - ¿Cree usted que existen diferencias entre cómo se percibe usted como profesor y cómo le perciben los demás (alumnos y compañeros)?

9. **Formación pedagógica**
   - ¿Cómo describiría la evolución que ha seguido como profesor desde que comenzó a dar clases hasta ahora?
   - ¿Qué es lo que más le ha ayudado a evolucionar en este sentido?
   - ¿Qué recomendaciones le haría a un profesor novel para que mejore su docencia y de qué obstáculos le prevendría?
   - ¿Participa o ha participado en alguna actividad formativa de carácter pedagógico?
10. **Necesidades docentes**
   - ¿Cómo sería para usted una organización ideal de la facultad y del departamento?
   - ¿Qué cambiaría o modificaría?

ANEXO II:
GUÍA TÉCNICA DE OBSERVACIÓN

Tipo:

Observación participante (participación moderada = Interacción limitada = Observador como participante)

Objetivos:

— Describir el proceso de enseñanza-aprendizaje en las aulas universitarias
— Describir las relaciones profesor-alumno
— Describir las relaciones del alumnado en el aula
— Describir el contexto de la enseñanza (tanto el contexto físico como la cultura que está implícita)

Ámbitos:

Clases teóricas y clases prácticas

Duración:

Tres sesiones de clase

Registro:

Notas de campo

Aspectos a observar:

1. **Aspecto físico del aula**:
   Localización del aula en el edificio, mobiliario, luminosidad del aula, tamaño, disposición del alumnado, etc.
2. **Metodología de enseñanza:**
   Incluye las estrategias didácticas que el profesorado emplea en sus clases:
   — lección, trabajo en grupo ... etc.
   — perfil típico de una clase
1. **Recursos**
   Materiales que se utilizan en las clases, su disposición y procedimiento de uso:
   — libros, apuntes, transparencias, retroproyector, vídeo, micrófono ... etc.
1. **Relaciones**
   Quiénes y cómo se relacionan en el aula
2. **Organización**
   Distribución del espacio y el tiempo durante las clases
3. **Evaluación**
   Actuaciones realizadas con la intención de valoración del rendimiento de los alumnos y la propia actividad docente
4. **Sentimientos**
   Cómo se siente el alumnado y el profesorado en la clase

ANEXO III:
GUIÓN DE ENTREVISTA SEMIESTRUCTURADA AL ALUMNADO

1. **Cumplimiento de las obligaciones por parte del profesor**
   - ¿Qué pensáis de una forma general sobre este profesor?
   - ¿Asiste regularmente a clase, justifica sus faltas, es puntual ... etc.?
   - ¿Cómo organiza las tutorías y la atención hacia vosotros? ¿Qué contenidos tratáis en ellas?
   - ¿Es abierto con vosotros y manifiesta interés por ayudaros?
2. **Información al comienzo del curso**
   - ¿Qué aspectos trató al comenzar la asignatura: su forma de dar las clases, la evaluación ...?
   - ¿Cómo os presentó estos aspectos? (Retroproyector, fotocopias ...etc)
   - ¿A qué dio más importancia en su asignatura (asistencia, ampliación, exámenes ...)?
   - ¿Os ha facilitado bibliografía complementaria?
   - ¿Os ha dado la oportunidad de discutir algunos aspectos del programa y de la marcha del curso?
3. **Relaciones del profesor con los alumnos**
   - ¿Podríais describir el modo en que se relaciona con vosotros?
   - ¿Cómo creéis que os ve a vosotros?
   - ¿Soléis participar en clase preguntando, criticando ...etc.?
   - ¿Cómo calificaríais a este profesor?
   - ¿Cómo describiríais el ambiente de clase?
4. **Sobre el sistema de evaluación**
   - ¿Qué opinión os merece su sistema de evaluación? ¿Qué pensáis sobre el mismo? ¿Participáis de alguna forma en la evaluación?
   - ¿Consideráis que el sistema es adecuado? ¿Qué cambiaríais del mismo?
5. **Sobre el desarrollo de las clases**
   - ¿Cómo describiríais una sesión de clase típica de este profesor?
   - ¿Qué opinión os merece la forma en que da clases?
   - ¿Utiliza algunos materiales en clase (retro, videos ...etc.)?
   - ¿Cómo calificaríais las clases? (Aburridas, interesantes ... etc.)
   - ¿Se busca la relación con otras asignaturas?
   - ¿Se para en los aspectos más importantes?
6. **Método de enseñanza**
   - ¿Cómo describiríais su forma de dar clase?
   - ¿Consideráis su forma de dar clase adecuada e interesante?
   - ¿Qué cambiaríais de sus clases?
7. **Opinión global**
   - ¿Estáis satisfecho en general con este profesor?

ANEXO IV:
SISTEMA DE CATEGORÍAS

1. Trayectoria profesional
   1.1. Continuidad en la materia
   1.2. Carga docente actual
   1.3. Acceso
   1.4. Experiencia
   1.5. Formación previa
2. Actividad profesional
   2.1. Docencia
      2.1.1. Relación docencia-investigación-gestión
   2.2. Investigación
   2.3. Gestión
   2.4. Otras: seminarios, cursos, revistas, tesis, etc.
3. Planificación de la enseñanza
   3.1. Actividades previas
   3.2. Coordinación con otros profesores
   3.3. Participación de los alumnos
   3.4. Diferencias entre teoría y prácticas
   3.5. Objetivos que se pretenden con la asignatura
   3.6. Selección de contenidos
4. Organización de la clase
   4.1. Actividades
   4.2. Tiempo
   4.3. Recursos
   4.4. Espacio
   4.5. Agrupamiento de alumnos
   4.6. Participación de los alumnos
   4.7. Estructuración de los contenidos
5. Evaluación
   5.1. Alumnos
   5.2. Autoevaluación
   5.3. Finalidad de la evaluación
6. Tutoría
   6.1. Formato
   6.2. Contenido
7. Relaciones
   7.1. Alumnos
   7.2. Profesores
8. Formación pedagógica
   8.1. Referentes
   8.2. Participación e iniciativas
   8.3. Orientaciones a noveles

9. Necesidades docentes
10. Estrategias de enseñanza
11. Cumplimiento de obligaciones
    - 11.1. Puntualidad
        - 11.1.1. Alumnos
        - 11.1.2. Profesores
    - 11.2. Asistencia
    - 11.3. Otros
12. Contexto: Sonoridad, tamaño, ubicación, luminosidad, etc.
13. Comunicación con el alumnado
    - 13.1. Verbal
    - 13.2. No verbal
14. Comunicación del alumnado con el profesor
15. Interés del alumnado por la asignatura
16. Alumnado de la clase
17. Formación general del profesor

424

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 433-445

LA EVALUACIÓN DE LA ENSEÑANZA POR LOS ALUMNOS EN EL PLAN NACIONAL DE EVALUACIÓN DE LA CALIDAD DE LAS UNIVERSIDADES. CONSTRUCCIÓN DE UN INSTRUMENTO DE EVALUACIÓN

Juan Ruiz Carrascosa
Universidad de Jaén

RESUMEN

En este artículo se considera la evaluación de la enseñanza en el ámbito de la evaluación de la calidad de la universidad y la aportación que pueden hacer los alumnos en este proceso. Se presenta un cuestionario construido para valorar las asignaturas que conforman el plan de estudios y los resultados del análisis multivariante realizado. El cuestionario se propone como un medio más para ser utilizado en la evaluación de la enseñanza dentro del Plan Nacional de Evaluación de la Calidad de las Universidades.

ABSTRACT

This article deals with the evaluation of teaching in the area of the evaluation of the quality of the university and what the students can add to this process. A questionnaire is presented in order to assess both the subjects that make up the curriculum and the results of the multivariant analysis carried out. The questionnaire is proposed as a further means in order to be used in the evaluation of teaching within the National Plan of Evaluation of the Quality of the Universities.

1. INTRODUCCIÓN

En los últimos años el debate público sobre la calidad en las universidades europeas viene siendo constante y son muchos los países, entre ellos España, que han estableci-

do mecanismos de evaluación de las universidades. La preocupación reciente por la búsqueda y el control de la calidad en la enseñanza universitaria se ha manifestado con fuerza en los años noventa, siendo un exponente de ello el desarrollo de procedimientos de evaluación de las instituciones de educación superior.

Entendemos que esta preocupación por la calidad deriva en parte de una exigencia de la propia sociedad. La universidad presta un servicio de interés general que es financiado, en la mayoría de los casos, principalmente con fondos públicos y es por ello que cada vez existe una mayor presión social sobre la calidad del servicio. Pero si bien es cierto que la presión externa ha sido un factor condicionante en el desarrollo de lo que podíamos denominar como la «*cultura de la calidad*» en la universidad, no podemos olvidar el peso que está teniendo en la expansión de esta cultura el interés creciente de los propios agentes internos de la institución universitaria por la calidad de los servicios que prestan y reciben y por la mejora de los mismos.

Una consecuencia de este interés creciente por la calidad es la preocupación generalizada por su evaluación, entendiendo la evaluación como un medio para la mejora de las instituciones universitarias (Consejo de la Unión Europea, 1998; Consejo de Universidades, 1999; Junta de Andalucía, 1999).

Con objeto de desarrollar sistemas de evaluación de calidad, durante los últimos quince años, la mayor parte de los países europeos han ido creando organismos responsables de evaluación de las universidades, cuya característica común es una independencia relativa, no sólo de los ministerios de educación sino también de las propias universidades. Estas innovaciones institucionales se inscriben en un movimiento general de descentralización de la gestión universitaria, variando mucho el estatuto jurídico y la composición de estos organismos en los diferentes países (European Commission. DGXXII, 1998).

Esta variedad de organización institucional relativa a la evaluación refleja, sin duda, las diferencias tradicionales que siguen persistiendo entre países con una gestión centralizada y aquellos en los que las universidades tienen una tradición de autonomía más antigua.

En España se confían las tareas de evaluación al Consejo de Universidades, desde el que se coordina el Plan Nacional de Evaluación de la Calidad de la Universidades (PNECU). En nuestro país, al igual que en otros países europeos como Holanda, Francia, Reino Unido y Dinamarca, que poseen sistemas consolidados de evaluación de sus Instituciones de Educación Superior, la metodología de evaluación empleada combina la autoevaluación (*self-evaluation*) con la evaluación externa (*peer review*). El modelo de evaluación que se aplica en España dentro del PNECU es el denominado «*de autorregulación*», que implica tres etapas: autoestudio, evaluación externa e informe final interno. La autoevaluación va seguida de la evaluación externa que proporciona credibilidad al proceso, valida las conclusiones (puntos fuertes / débiles y propuestas de mejora) y, en definitiva, refuerza la garantía de calidad.

En este modelo «*de autorregulación*» el PNECU prevé que la unidad de evaluación se referencie en la titulación, evaluándose al tiempo todas las actividades de la universidad: enseñanza, investigación y servicios (figura 1).

426

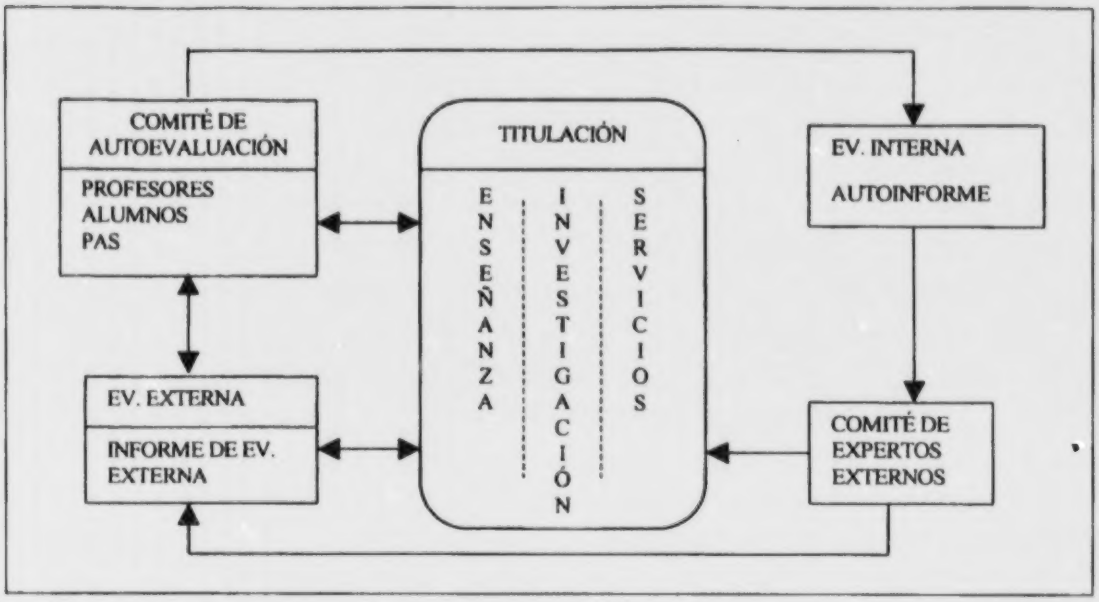


Figura 1
Evaluación de una titulación en el PNECU

2. LA EVALUACIÓN DE LA ENSEÑANZA

El logro de una enseñanza de calidad en la educación superior se nos ofrece hoy como un objetivo irrenunciable y de carácter prioritario. Las recomendaciones en este sentido son cada vez más frecuentes, destacando la necesidad de atender en mayor medida la enseñanza, a veces desfavorecida por una atención prioritaria a las actividades de investigación (Informe Dearing, 1997; Consejo de la Comisión Europea, 1998; Consejo de Universidades, 1999).

En los últimos cinco años el desarrollo del PNECU ha puesto de manifiesto un interés creciente en el ámbito universitario por la evaluación de la enseñanza, reflejándose en los diferentes informes la atención otorgada a esta dimensión en la evaluación de las titulaciones (Consejo de Universidades, 2000).

La evaluación de la enseñanza (education) como destaca Rodríguez (1999) ha llegado, en algunas ocasiones, desde planteamientos reduccionistas a identificarse con los estudios de la calidad de la docencia (Teaching). En la actual práctica de la evaluación institucional de la enseñanza, orientada a la mejora, esta identificación no es pertinente. Entendemos que cualquier esfuerzo de evaluación de las dimensiones consideradas en el ámbito de la enseñanza ha de integrarse, desde un enfoque global, con los datos aportados en el resto de dimensiones y aspectos considerados.

En la evaluación de la enseñanza, dentro del PNECU se consideran todas aquellas áreas que tienen relación con la titulación: Contexto, Metas y Objetivos, Propuestas de

Formación, Desarrollo de la Enseñanza, Estudiantes, Profesores, Instalaciones y Recursos, así como Relaciones Externas. En el análisis que se realiza de todas estas áreas durante el proceso de autoevaluación, en el que participan profesores, alumnos y Personal de Administración y Servicios (PAS), una parte fundamental es la recogida y sistematización de la información para su análisis y valoración posterior. La calidad de la información recogida así como la atención prestada a este proceso va a condicionar la elaboración y calidad del propio autoinforme.

3. LA RECOGIDA DE INFORMACIÓN

Considerando que en la evaluación han de obtenerse evidencias (información objetiva de índole cuantitativa y cualitativa) de modo sistemático para informar los juicios de valor consustanciales a todo proceso evaluativo permitiendo la toma de decisiones y teniendo en cuenta la diversidad de áreas que se abordan, las fuentes de información han de ser variadas así como los instrumentos utilizados para su recogida.

Las fuentes de información básicamente pueden agruparse en dos grandes grupos, en primer lugar aquellas que aportan datos objetivos provenientes de las estadísticas universitarias (ratios, calificaciones, tasas de éxito, fracaso y abandono, espacios y medios disponibles, etc...). En segundo lugar se encuentran las que aportan datos basados en la opinión, pero en los que pueden asegurarse ciertos niveles de validez y fiabilidad. En este caso los datos se obtienen fundamentalmente mediante encuestas, entrevistas, observación y grupos de debate.

La obtención de datos de las primeras fuentes es todavía un problema importante en los procesos de evaluación institucional. En el segundo grupo, hay que avanzar en la búsqueda de información objetiva a través de los diferentes medios e instrumentos utilizados.

Las aportaciones de los distintos agentes implicados en la evaluación de una titulación profesores, alumnos y PAS resultan muy importantes en este sentido. La información de los alumnos a través de la valoración de las asignaturas que conforman el plan de estudios puede aportar datos de gran valor en el proceso de evaluación de una titulación. Por ello la recogida de información a través de la opinión de los alumnos ha de hacerse mediante instrumentos y procedimientos que garanticen el máximo de validez y fiabilidad a los datos obtenidos.

4. EL INSTRUMENTO DE RECOGIDA DE INFORMACIÓN SOBRE LA OPINIÓN DE LOS ESTUDIANTES

A través de una encuesta de opinión a los estudiantes sobre las asignaturas de la titulación en que cursan sus estudios, mediante la aplicación de un cuestionario, se pretende obtener información válida y fiable acerca de las variables consideradas en las distintas dimensiones contempladas, para su integración en el autoinforme y la correspondiente toma de decisiones.

El cuestionario ha sido aplicado, en una experiencia piloto, a una muestra de 434 alumnos de las titulaciones evaluadas dentro del PNECU en la universidad de Jaén.

La información obtenida a través de este cuestionario sobre la valoración por parte de los alumnos de las asignaturas que conforman el plan de estudios de una titulación puede ser útil combinada con aquellos datos que se contemplan en el autoinforme (tablas 4 y 5 de la Guía del PNECU) sobre los programas de las asignaturas, tiempo de estudio semanal, estrategias docentes utilizadas y tipología, organización y desarrollo de las prácticas.

4.1. Validez y fiabilidad del cuestionario

En la construcción del cuestionario se ha procurado garantizar la validez y fiabilidad del mismo a través de diversos procedimientos. La fiabilidad, obtenida mediante el α de Chronbach, ha dado un valor igual a 0,9871, por lo que puede considerarse como muy elevada. En cuanto a la validez, en primer lugar se ha efectuado la correspondiente revisión bibliográfica del tema teniendo como referente principal la Guía de Evaluación del PNECU, para determinar las variables que debieran ser consideradas sometiéndose éstas a una valoración de jueces expertos.

Finalmente se ha efectuado un análisis factorial con objeto de identificar los factores que puedan ser utilizados para representar las relaciones entre las variables consideradas.

4.1.1. Análisis factorial

En la figura 2 aparecen en primer lugar la media y la desviación estándar de cada una de las 34 variables entradas en el análisis factorial a las que han respondido los 434 sujetos de la muestra a través de las preguntas del cuestionario a las que se refieren estas variables.

	Media	Desv. típica
P1	3.35	1.17
P2	3.61	1.06
P3	3.10	1.14
P4	3.10	1.28
P5	2.97	1.37
P6	3.00	1.19
P7	2.94	1.19
P8	3.30	1.08
P9	3.24	1.23
P10	2.89	1.29
P11	3.12	1.15
P12	4.39	.87
P13	4.09	1.12
P14	3.40	1.35
P15	2.84	1.28
P16	3.04	1.45
P17	3.02	1.43

	Media	Desv. típica
P18	2.74	1.40
P19	3.33	1.36
P20	3.03	1.39
P21	3.02	1.37
P22	3.07	1.31
P23	2.99	1.34
P24	2.93	1.37
P25	3.03	1.39
P26	3.59	1.25
P27	3.76	1.07
P28	3.20	1.31
P29	3.50	1.22
P30	3.38	1.25
P31	3.47	1.31
P32	3.28	1.26
P33	3.49	1.13
P34	3.78	1.12

Figura 2
Estadísticos Descriptivos

El grado de significación de los coeficientes que aparecen en la matriz de correlaciones entre variables, en un contraste unilateral, resulta significativo en todos los casos $p \leq 0{,}001$. El determinante de la matriz de correlaciones es 1,772E-1.

En la figura 3 se ofrecen los KMO (Káiser-Meyer-Olkín) y el test de Bartlett.

Sí la suma de los coeficientes de correlación parcial al cuadrado, entre variables, es muy pequeña, KMO será un índice muy próximo a la unidad y por tanto el análisis factorial un procedimiento adecuado (Visauta, 1998). En nuestro caso este valor es de 0,962 y se puede considerar como excelente, por lo que procede el análisis factorial.

El test de Bartlett se utiliza para verificar si la matriz de correlaciones es una matriz de identidad, es decir si todos los coeficientes de la diagonal son iguales a la unidad y los extremos a la diagonal iguales a 0. Este estadístico, obtenido a partir de la transformación Chi-cuadrado del determinante de la matriz de correlaciones, cuanto mayor sea y por tanto menor el grado de significación es más improbable que la matriz sea una matriz de identidad. En nuestro caso con un valor 13173,490 y un grado de significación $p = 0{,}000$ resulta evidente que no se trata de una matriz de identidad. Esto nos lleva a afirmar que el análisis factorial que sigue a continuación resulta pertinente y puede proporcionar conclusiones satisfactorias.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.962
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	13173.490
	gl	561
	Sig.	.000

Figura 3
KMO y prueba de Bartlett

4.1.1.1. Extracción de Factores

El método utilizado es el de componentes principales, cuyo objetivo consiste en encontrar una serie de componentes que expliquen el máximo de varianza total de las variables originales.

En la figura 4 vemos que hay cinco factores con valores propios superiores a 1 y que en definitiva será el número que extraerá el sistema. Igualmente, en esta figura se recoge, en porcentajes individuales y acumulados, la proporción de varianza total explicada por cada factor, tanto para la solución no rotada como para la rotada. Los cinco factores incluidos en el modelo son capaces de explicar exactamente un 72,276 por 100 de la variabilidad total.

	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
Componente	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	18,243	53.655	53.655	18,243	53.655	53.655	9.304	27.365	27.365
2	1.941	5.710	59.365	1.941	5.710	59.365	5.196	15.282	42.647
3	1.814	5.337	64.702	1.814	5.337	64.702	4.828	14.199	56.846
4	1.342	3.947	68.649	1.342	3.947	68.649	2.636	7.753	64.599
5	1.233	3.627	72.276	1.233	3.627	72.276	2.610	7.677	72.276
6	.972	2.858	75.134						

Método de extracción: Análisis de Componentes principales.

Figura 4
Estadísticos iniciales

En la figura 5 tenemos una representación gráfica de estos resultados, figurando en el eje de abcisas el número total de factores y en ordenadas el valor propio de cada uno de ellos.

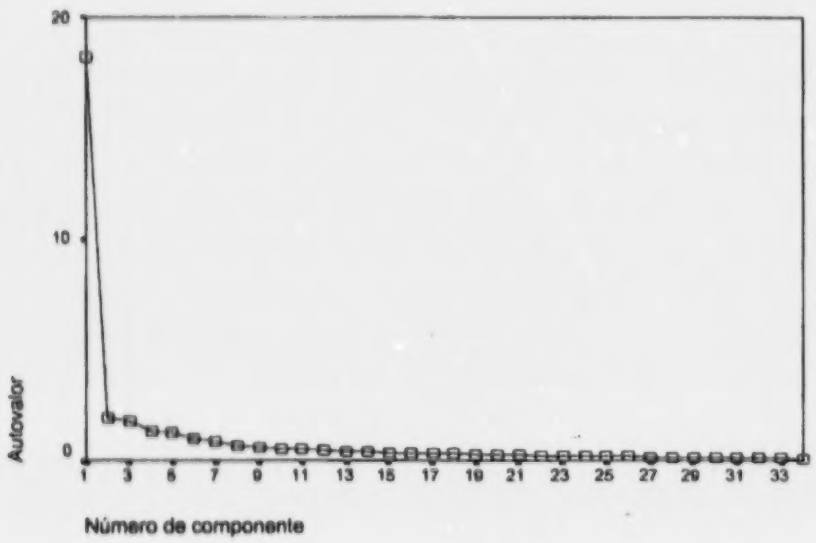


Figura 5
Gráfico de sedimentación

Utilizando el método de componentes principales y rotación VARIMAX obtenemos la matriz de la figura 6.

Matriz de componentes rotados [a]

	Componente				
	1	2	3	4	5
p24	.837				
p23	.824				
p20	.811				
p25	.796				
p21	.775				
p22	.772				
p28	.743				
p19	.722				
p14	.714				
p18	.694				
p17	.670				
p16	.635				
p26	.594				
p7		.750			
p10		.748			
p6		.686			
p5		.679			
p4		.645			
p15		.585			
p3		.584			
p9		.524			
p11		.501			
p29			.778		
p30			.767		
p31			.682		
p33			.614		
p32			.603		
p34			.572		
p2				.789	
p8				.596	
p1				.470	
p12					.801
p13					.748
p27					.580

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 8 iteraciones.

Figura 6
Matriz de pesos factoriales y rotación VARIMAX

Los cinco factores obtenidos se han denominado del siguiente modo:

Factor 1: Actuación docente
Factor 2: Organización y medios
Factor 3: Evaluación
Factor 4: Planificación de la asignatura
Factor 5: Cumplimiento formal del profesor

En la figura 7 aparecen las variables agrupadas en cada uno de los cinco factores obtenidos en el análisis.

Factor 1: Actuación docente
- P24.- Actuación del profesor para crear interés por la asignatura
- P23.- Potenciación del diálogo
- P20.- Estimula la participación del alumno en clase
- P25.- El profesor procura indagar si los estudiantes entienden lo que explica
- P21.- Estimula el trabajo personal de los estudiantes en la asignatura
- P22.- Se interesa por los alumnos
- P28.- Cómo valoras globalmente la actuación docente del profesor
- P19.- Responde con exactitud y precisión a las cuestiones que se le plantean
- P14.- Motivación del profesor para la docencia
- P18.- Amenidad
- P17.- Eficacia en la transmisión de conceptos
- P16.- Claridad y orden en la presentación de la materia
- P26.- Disponibilidad del profesor para consultas, sugerencias, etc.

Factor 2: Organización y medios
- P7.- Condiciones materiales de las clases de prácticas
- P10.- Disponibilidad de materiales de apoyo a la docencia (transparencias, esquemas...)
- P6.- Condiciones materiales de las clases de teoría
- P5.- Organización de las prácticas
- P4.- Coordinación entre teoría y resto de actividades
- P15.- Utilización de medios y recursos didácticos
- P3.- Material recomendado por el profesor
- P9.- Información bibliográfica para la preparación del programa de la asignatura
- P11.- Accesibilidad en biblioteca de la bibliografía y lecturas recomendadas por el profesor

Factor 3: Evaluación
- P29.- Información adecuada, con tiempo suficiente, de cómo sería el sistema de evaluación utilizado en esta asignatura
- P30.- Precisión en la definición de criterios de evaluación
- P31.- Correspondencia del nivel de los exámenes con lo tratado en clase
- P33.- Tiempo transcurrido entre la realización de un examen y la comunicación del resultado
- P32.- Relación entre los resultados obtenidos en los exámenes y los conocimientos adquiridos
- P34.- Accesibilidad para la revisión del examen

Factor 4: Planificación de la asignatura
- P2.- Importancia de los contenidos de la asignatura para la formación
- P8.- Coordinación de la asignatura con el resto de la carrera
- P1.- Definición de los objetivos y contenidos de la asignatura

Factor 5: Cumplimiento formal del profesor
- P12.- Asistencia a clase
- P13.- Puntualidad
- P27.- Disponibilidad del profesor en su horario de tutorías

Figura 7
Variables contempladas en cada factor

5. ANÁLISIS DE REGRESIÓN

El cuestionario, junto a las 34 preguntas que lo conforman, presenta una pregunta de valoración global de la asignatura, por lo que hemos efectuado un análisis de regresión múltiple. Se ha ajustado un modelo que explica la nota otorgada a la valoración global de la asignatura en función de las otorgadas a los diferentes factores. En la figura 8 se contempla un resumen del modelo, donde se han considerado variables predictoras los cinco factores y variable dependiente la valoración global de la asignatura.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error tip. de la estimación
1	.882[a]	.779	.776	.65

a. Variables predictoras: (Constante), FACTOR5, FACTOR4, FACTOR3, FACTOR2, FACTOR1.

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	621.488	5	124.298	291.970	.000[a]
	Residual	176.674	415	.426		
	Total	798.162	420			

a. Variables predictoras: (Constante), FACTOR5, FACTOR4, FACTOR3, FACTOR2, FACTOR1.
b. Variable dependiente: p. 35.

Figura 8
Resumen del modelo y ANOVA

A continuación se ofrecen los coeficientes de regresión obtenidos (figura 9) y la ecuación de regresión.

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error tip.	Beta		
1 (Constante)	-.522	.175		-2.975	.003
FACTOR1	.676	.050	.582	13.602	.000
FACTOR2	-7,776E-02	.056	-.055	-1.389	.166
FACTOR3	.441	.048	,333	9.206	.000
FACTOR4	.238	.051	.154	4.693	.000
FACTOR5	-.132	.048	-.079	-2.755	.006

a. Variable dependiente: p. 35.

Figura 9
Coeficientes de regresión

La ecuación es:

VALORACIÓN GLOBAL = -0,552 + 0,676*ACTUACIÓN DOCENTE - 7.776E-02*ORGANIZACIÓN Y MEDIOS + 0,441*EVALUACIÓN + 0,238*PLANIFICACIÓN DE LA ASIGNATURA - 0,132*CUMPLIMIENTO FORMAL DEL PROFESOR

Esta ecuación indica que la actuación docente del profesor es la variable que predice en mayor medida la valoración global de una asignatura por parte de los alumnos, haciéndolo a continuación la evaluación y la planificación de la asignatura.

Para fortalecer esta consideración mostramos en la figura 10 el comportamiento de los datos analizados frente a los datos predichos por el modelo. Como se puede apreciar, la concordancia entre lo que predice el modelo y la valoración de los alumnos es bastante elevada.

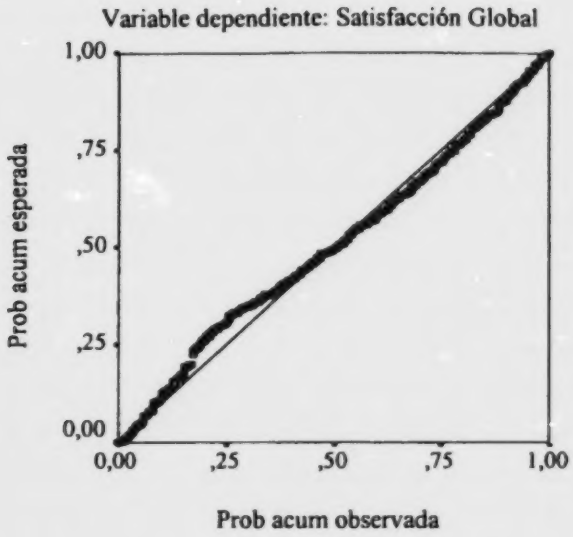


Figura 10
Gráfico P-P normal de regresión

6. CONCLUSIONES

La construcción del cuestionario para recoger la valoración de los estudiantes sobre las asignaturas de una titulación ha permitido disponer de un instrumento válido y fiable para la obtención de información en el proceso de evaluación de la enseñanza, dentro del PNECU.

Los datos obtenidos a través de las variables que integran los cinco factores aportan una información útil en el proceso de avaluación institucional de la enseñanza. No obstante, los datos obtenidos a través de instrumentos de este tipo no deberían considerarse como única fuente de información respecto a los estudiantes.

En la evaluación de la calidad de la enseñanza hay que destacar la necesaria complementariedad en las distintas técnicas e instrumentos utilizados. En el caso concreto de las asignaturas de una determinada titulación, la valoración que efectúan los alumnos de cada una de ellas deberá completarse con la información que se contempla en el autoinforme sobre el programa de las asignaturas y las prácticas en el plan de estudios.

Finalmente hay que destacar la necesidad de avanzar en la integración de los datos obtenidos en el proceso de evaluación de la enseñanza y en el análisis de sus relaciones. Hay que analizar en que medida interaccionan factores como: objetivos específicos de los programas, métodos de enseñanza empleados, relación entre tipología de evaluación del alumnado y tipo de aprendizaje, resultados en términos de conocimientos y actitudes, y todo ello considerando las condiciones del marco de referencia propio.

7. BIBLIOGRAFÍA

Consejo de la Unión Europea (1998). *Recomendación del Consejo sobre la cooperación europea para la garantía de la calidad en la enseñanza superior.* (DOC 7, 10, 1998).

Consejo de Universidades (1999). *Plan Nacional de Evaluación de la Calidad de las Universidades.* Madrid: Consejo de Universidades. Secretaría General.

Consejo de Universidades (2000). *Informe Final. Segunda convocatoria 1998.* Madrid: MEC. Consejo de Universidades.

European Commitssion, DGXII (1998). *Evaluation of European Higher Education: A Status Report.*

Junta de Andalucía (1999). *Unidad para la Calidad de las Universidades Andaluzas.* Sevilla: Consejería de Educación y Ciencia.

Rodríguez, S (1999). La Investigación en Evaluación Institucional. *Cuadernos IRC.* Nº 3. Pp. 31-38.

The National Committee of Inquiry Into Higher Education (1997). *Higher Education in the learning society. Report of the National Committee.* Great Britain.

Visauta (1998). *Análisis Estadístico con SPSS para Windows.* Madrid: McGraw-Hill.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 447-461

EVALUACIONES DE PROGRAMAS DE FORMACIÓN UNIVERSITARIA EN EL ÁMBITO EUROPEO Y AMERICANO. COHERENCIA CON LAS DEMANDAS SOCIALES

Honorio Salmerón Pérez

Dpto. de Métodos de Investigación y Diagnóstico en Educación
Universidad de Granada

1. INTRODUCCIÓN

La evaluación se ha constituido progresivamente como una lógica que va más allá de una simple tecnología de medición y control. Se plantea como una herramienta para esclarecer el trabajo colectivo, y para crear una comunidad universitaria crítica y autocrítica, capaz de hacer de su autonomía una afirmación de identidad social. La evaluación es creadora de comunidad, porque debe permitir construir una reflexión común no sólo sobre la realización concreta de objetivos, sino también sobre las finalidades que los animan, que son las que dan sentido al quehacer colectivo de creación de conocimientos y de significaciones. Para hablar de creación de sentido colectivo es necesario hablar de proyecto colectivo, de proyecto de universidad. Ello implica contextualizar la evaluación en los fines de la institución que ella observa, juzga y orienta.

ZÚÑIGA, R. (1994. Universidad de Montreal, Quebec, Canadá) considera, que el desafío actual está en cómo evaluar la calidad en contextos en los que no están claros los objetivos o metas a alcanzar, a lo que denomina «desorientación universitaria». El suponer que las universidades saben hacia dónde van, reflexionan dinámicamente sobre su misión, para asegurarse de que se insertan en un momento histórico, en una sociedad dada, y que lo hacen de una forma que les permita derivar unas metas concretas y planes de acción, es una suposición que peca de generosa.

Esta desorientación se expresa en varios rasgos:

— En primer lugar, la autonomía real de la universidad merece un análisis crítico. Está cada vez mas claro que las fuentes de financiación son y serán cada vez más fuentes múltiples. Los gobiernos reducen progresivamente sus aportes e instauran sutilmente el discurso de «la universidad abierta», una universidad que escucha las voces de la comunidad, y sobre todo las del mercado, para definir su misión.
— En segundo lugar, para la evaluación de la calidad global, hace ya tiempo que hemos subordinado la evaluación interna a la evaluación como imagen pública, a los esfuerzos de nuestros administradores para asegurar la contribución económica del Estado.
— La universidad es cada vez más percibida vista como cualquier otra institución, y los privilegios que reivindica son cada vez más vistos como arcaísmos.
— El papel central que juega y puede jugar la evaluación en el quehacer universitario sería:
  — Como instrumento técnico de control interno, de responsabilidad y de imputabilidad.
  — Como instrumento de formación. Formación de intelectuales, de profesionales; formación de comunidad, de respuesta a una misión de sociedad.

La filosofía que emerge de la experiencia acreditadora americana, es contrastante con el panorama europeo. El continente europeo es heredero de la tradición centralista en materia de planeamiento educativo (sobre todo a partir de los 70), cuya expresión en el ámbito universitario significó la intervención del Estado en los asuntos académicos. Esta tendencia difiere de los procesos de autorregulación norteamericanos, por considerar a la institución universitaria como la única y legítima responsable del proyecto académico.

Las universidades pueden proponerse un perfil de calidad respetando la tradición y la propia dinámica institucional; no es necesario que las otras universidades del sistema (laicas o confesionales, públicas o privadas) alcancen el mismo resultado, porque pueden partir de otro tipo de definición de calidad. Lo que es fundamental es la coherencia entre la meta que cada universidad se propone alcanzar, los medios y los resultados obtenidos. Pero, ¿están claras esas metas?, ¿hay una meta o hay varias?, si son varias, ¿se contradicen entre sí?, ¿existen los recursos suficientes para conquistar dichas metas?, la formación de profesionales críticos, la formación de graduados con capacidad y habilidad para resolver su práctica profesional e insertarse en el medio laboral, la formación de investigadores, etc. ¿Constituyen todas ellas metas compatibles?

En este momento nuestra preocupación la centramos en la inserción laboral de titulados con competencias profesionales homologadas con sus colegas de la Unión Europea. En este sentido resaltamos dos disposiciones del Consejo de Europa al respecto:

Resolución del Consejo, de 5 de diciembre de 1994, sobre la calidad y atractivo de la Formación Profesional.
Diario Oficial nº C 374 de 30/12/1994 P. 0001- 0004.

«...la política de formación profesional debería contribuir de manera importante a hacer efectiva la libre circulación y a estimular la movilidad de los trabajadores dentro del mercado interior europeo. No se trata únicamente de trasmitir —competencias europeas—, sino asimismo de permitir que los certificados y títulos profesionales nacionales puedan utilizarse más allá de las fronteras en el mercado europeo del trabajo».

*** Recomendación del Consejo de 24 de septiembre de 1998 sobre la cooperación europea para la garantía de la calidad en la enseñanza superior.**
Diario Oficial nº L 270 de 07/10/1998 P. 0056-0059

«Se pide que la Comunidad contribuya a ese esfuerzo continuo mediante el fomento de la cooperación entre los Estados miembros y apoyando y completando su acción, si fuere necesario, y respetando plenamente, al mismo tiempo, la responsabilidad de estos últimos en lo que se refiere al contenido de la enseñanza y a la organización de los sistemas educativos y de formación, así como su diversidad cultural y lingüística».

2. LA EVALUACIÓN DE PROGRAMAS DE FORMACIÓN UNIVERSITARIA

La condición fundamental de una institución universitaria es la de su calidad y su compromiso de mejorarla de manera constante. Cuando estas dos condiciones están garantizadas, se da una real presencia en el tejido social a través de los soportes de la docencia, investigación y extensión.

Como señala ASTIN (1996), la evaluación es una herramienta potencialmente poderosa para ayudarnos a construir un programa educativo más eficaz y eficiente ya que nos aporta estrategias para mejorar el proceso de enseñanza-aprendizaje al ponernos de manifiesto las potencialidades y debilidades del programa de formación.

La planificación, implementación y evaluación de un programa universitario implica la participación activa y coordinada de grupos de personas que deben desarrollar tareas complejas interdependientes. La comunicación, participación y negociación son cuestiones fundamentales para el éxito del programa. Desde el punto de vista de la psicología organizacional, problemas de motivación, confianza, conflicto o comunicación pueden derivar en el éxito o fracaso del programa (REBOLLOSO Y MORTALE, 1996).

Cuando hablamos de «evaluación de programas de formación», nos estamos refiriendo al mismo tiempo a la efectividad y eficiencia de los planes de estudio. El resultado de la evaluación, en este caso de programas de formación universitaria, es un informe, a partir del cual se diseñarán las nuevas líneas estratégicas a seguir, las propuestas de mejora.

La mayoría de los informes de evaluación, señalan, a parte de los aspectos intrínsecos de cada institución evaluada, otras cuestiones de carácter global entre las que destacamos las siguientes:

— Promover una cultura sobre la función docente como un servicio social que exige a las instituciones y al profesorado plantear, enfocar y adaptar su trabajo

desde la perspectiva de las necesidades sociales al margen de sus intereses académicos personales.

SOLIMAN (1984) encontró en sus estudios realizados en la Universidad de Australia, que los profesores valoraban de la misma manera la investigación que la enseñanza, aunque reconocían que la investigación era una vía más fácil de promoción. Once años más tarde, MASSY & WILGER,1995, demuestran que el personal académico considera mucho más importante su función investigadora que su función docente. Del mismo modo, MOORE & GARDNER'S (1992) en la Universidad del Estado de Michigan, encuentran que la mayoría de los académicos dedican un 35% a la docencia, un 26% a la investigación y el resto en otras actividades, poniendo de manifiesto, su interés en reducir el tiempo invertido en docencia y actividades administrativas.

— Establecer mecanismos de seguimiento de los graduados en su proceso de inserción profesional que no sólo nos informe sobre las expectativas del mercado de trabajo sino que también actúe como realimentación de los perfiles académicos y profesionales de las titulaciones.
— Generar políticas de racionalización de los recursos materiales y humanos disponibles a fin de evitar las ineficiencias y disfuncionalidades observadas.
— Intensificar los convenios e intercambios de la titulación con el contexto social y laboral próximo a fin de lograr una mayor proyección y vinculación de la institución universitaria con la sociedad.

El Libro Blanco de la Comisión Europea sobre «crecimiento, competitividad y empleo» (1993) destacaba ya la «necesidad en la adquisición de competencias no especializadas, sino transversales y con cada vez mayor formación base». Del mismo modo, la Comunicación relativa a la evaluación de la aplicación del Libro blanco «Enseñar a aprender: hacia la sociedad cognitiva» (29 de mayo de 1997), pone de manifiesto entre otros aspectos:

— Se considera que cualquier mejora en la educación y la formación, contribuye a reforzar el crecimiento y la competitividad de Europa; análisis compartido por todos, especialmente por los representantes de los empresarios.
— Se reconoce ampliamente la necesidad de mejorar la convergencia entre enseñanza general y profesional, así como de valorar la formación profesional en su conjunto.

3. ANÁLISIS DE ALGUNAS REALIDADES: INFORMES DE EVALUACIÓN

A continuación exponemos algunos informes seleccionando, de éllos, la información que creemos más oportuna dado el propósito de nuestra aportación:

— Dentro de los resultados de la Primera Convocatoria del Plan Nacional de Evaluación de la Calidad de las Universidades llevado a cabo por el Consejo de Universidades de España, 1997, se pone ya de manifiesto la *disfuncionalidad de los planes*

de estudio:» Existe una gran disparidad de criterios en relación con la estructura del programa formativo de cada titulación que se pone de manifiesto por el desajuste entre la dimensión teórica y práctica de los planes, por la disparidad en los índices de optatividad, por la configuración de las materias de libre configuración, o por la falta de previsiones sobre los niveles formativos previos, etc...»

— En el mismo sentido, el estudio llevado a cabo por el Ministerio de Trabajo y Asuntos Sociales español y el Instituto Nacional de Empleo «El Mercado Laboral de los Titulados Universitarios», en 1996, que sigue una línea de continuidad —en sus aspectos principales— con otra publicación anterior del mismo título editada en 1994, ofrece datos y sugerencias para disminuir la preocupación, dentro del mundo universitario, sobre la capacidad real del mercado de trabajo para poder absorber la gran cantidad de estudiantes que cada año se gradúan. También sirve para poner de relieve el grado de adecuación que se está produciendo entre el sistema educativo —y sus políticas académicas— y el tejido empresarial. En este sentido, el estudio pone especialmente de manifiesto, *los desequilibrios o desajustes que se producen entre la formación académica que reciben los titulados y las exigencias de la oferta empresarial*

A continuación se exponen otros casos prácticos sobre informes de evaluación de formación universitaria:

- **«LA VISIÓN DE LOS DOCENTES SOBRE EL PLAN DE ESTUDIO»**
  Héctor Enrique Tamburini. FACULTAD DE INGENIERÍA, (1996). Universidad Nacional del Nordeste (Argentina).

— El documento constituye una de las ponencias referentes al «II Encuentro Interfacultades sobre Evaluación de la Calidad» celebrado durante el 29 de febrero y 1 de marzo en Resistencia (Argentina).

El análisis del Plan de Estudio de las carreras de ingeniería, indispensable en el Programa de Cambio Curricular, presenta los siguientes resultados:

PLAN DE ESTUDIO

	Responde a las necesidades y demandas actuales de la región y/o país?	
Carreras	**SÍ**	**NO**
Ingenie. En Construcción	76%	24%
Ingenie. Vías de Comunicación	61%	39%
Ingenie. Hidraúlica	50%	50%
Ingeniería Civil	65%	35%
Ingiene Mecánica	58%	42%
Ingenie. Electromecánica	81%	19%

En este cuadro podemos observar que el porcentaje afirmativo es aceptable, no obstante, en muchos casos, los profesores consideran excesivos los conocimientos en detrimento de otros importantes, teniendo en cuenta, los acelerados cambios tecnológicos.

En cuanto al Perfil Profesional del actual egresado, el resultado muestra lo siguiente:

	Perfil profesional del actual egresado					
	Demanda de profesionales en su especialidad			Formac. Académica insuf. y/o no actualizada según avances cientf.-técnicos		
Carreras	Suficiente	Insufic.	No responde	Sufi-ciente	Insufic.	No responde
Ineniería en Construcción	40%	43%	17%	65%	30%	5%
Ingenie. Vías de Comunicc.	35%	50%	15%	65%	30%	5%
Ingeniería Hidraúlica	30%	40%	30%	65%	30%	5%
Ingeniería Civil	35%	45%	20%	65%	30%	5%
Ingeniería Mecánica	30%	55%	15%	65%	30%	5%
Ingenie. Electromecánica	60%	30%	10%	65%	30%	5%

— Con respecto al perfil profesional del actual egresado, los docentes sostienen el mantenimiento de un perfil generalista, con fuerte formación básica con capacidad de ingenio para proyectar, construir y mantener todo tipo de obras civiles, percibir los cambios, en lo posible anticiparse a ellos, y con cierta formación humanística.

- **«INFORME DE EVALUACIÓN DE CALIDAD»**
  FACULTAD DE ARQUITECTURA. UNIVERSITY OF CAMBRIDGE, 1994.

Este informe refleja el resultado obtenido tras el proceso de evaluación de calidad llevado a cabo en la Facultad de Arquitectura de la Universidad de Cambridge cuyo resultado es «Excellent».

El diseño y estructura del plan de estudios cuenta con una amplia variación de métodos de enseñanza, incluyendo el uso de lugares y edificios reales. El volumen de conocimientos adquirido es amplio y muy completo, gracias a la labor, en profundidad, llevada a cabo por el Departamento encargado de las materias. Este intensivo programa de estudio está respaldado por las Escuelas de Arquitectura y por la avalada experiencia del personal involucrado.

Los programas de estudio están diseñados de tal manera que se logre en el estudiante un alto nivel de motivación y capacidad organizativa para llevar a cabo sus propios proyectos y

alcanzar sus objetivos durante la carrera. Para ello cuenta con una infraestructura de apoyo y asesoramiento permanente, no obstante, como consecuencia de la independencia en las unidades de estudio, los objetivos del proyecto de trabajo a alcanzar en las distintas etapas de la carrera del estudiante, no están claramente explícitos ni ampliamente desarrollados.

Conclusiones y recomendaciones

Los resultados, obtenidos en octubre de 1995 por el grupo de evaluadores de la Universidad de Cambridge, basados en el informe de evaluación externo, en el análisis interno y en indicadores estadísticos, revelan que la calidad en la educación recibida en la Facultad de Arquitectura de la Universidad de Cambridge, es «EXCELLENT».

Los objetivos a alcanzar en las distintas etapas del «proyecto de trabajo» de los estudiantes, no están muy definidos, no obstante, se cuenta ya con una amplia diversidad en la exploración de nuevas ideas y actitudes en el proyecto de trabajo.

- **«THE MACGILL UNIVERSITY VALUE ENGINEERING WORKSHOP» (Taller de Ingeniería de la Universidad Mcgill).**
  FACULTAD DE INGENIERÍA. UNIVERSIDAD McGILL (EEUU). 1994/95.INCE THOMSON.

El Departamento de Ingeniería Mecánica de la Universidad de Mcgill trabajó en un **curso práctico** «Value ingeneering» durante los pasados 22 años, donde la **cooperación industria-universidad** es única.

Este es el único curso universitario donde compañías representativas en el sector se sientan junto a estudiantes de ingeniería mecánica para resolver sus problemas en común.

La compañía expone un problema de la vida real que los estudiantes tratarán de resolver, a través de sus proyectos, de los cuales uno será seleccionado por la compañía para ser llevado a la práctica.

Las sesiones se llevaron a cabo durante cinco días a lo largo de cinco semanas, acumulando un total de 40 horas. Adicionalmente había una sesión práctica final para la presentación de los proyectos. El curso estaba tutelado por un profesor especialista y de avalada experiencia, perteneciente al Departamento de Ingeniería Mecánica, y por un grupo de expertos que comparten sus largos años de experiencia profesional con el grupo de estudiantes, aplicando los métodos de ingeniería en la resolución de problemas.

Comentarios del personal Participante:

«Nosotros llevamos a cabo el proceso de venta del proyecto durante cinco años. Esto nos convirtió en la mayor compañía en este tipo de negocios de Norte América». (Presidente, Air-dryin equipment manufacturer).

— «El mejor curso que he realizado. Realmente valoro la experiencia adquirida trabajando en el proyecto de industria». (estudiante que ha terminado el curso).
— «Me he divertido mucho trabajando con los estudiantes. Una gran experiencia. La sensación de que entra mucha energía en esto». (Personal especializado).

«EVALUACIÓN DE LOS PLANES DE ESTUDIO»
FACULTAD DE HUMANIDADES DE LA UNNE (Universidad Nacional del Nordeste. Argentina). Diciembre de 1995. Patricia Núñez de Uberti y Sonia Alejandra Zidianakis.

El presente documento constituye otra de las ponencias celebradas durante el «II Encuentro Interfacultades sobre Evaluación de la Calidad» en la ciudad de Resistencia (Argentina).

La Facultad de Humanidades, estuvo desarrollando un proceso de autoevaluación, que en esta ocasión tenía como tema de análisis, la evaluación de los planes de estudio. En este aspecto, los propósitos institucionales evaluados se referían a:

1. Necesidad de la carrera y pertinencia de los planes de estudio.
2. Calidad de la formación brindada a los alumnos.

Para evaluar los propósitos se analizaron las condiciones existentes e inexistentes, que se estimaban necesarias para el nivel de logro óptimo de los mismos.

El primer propósito era: «Ofrecer carreras que por su nivel y contenido, satisfagan reales necesidades emergentes de las demandas sociales, científicas, económicas y culturales de la región, del país y de los proyectos y políticas de desarrollo y crecimiento que se promuevan».

Para evaluar su nivel de logro se consideraron dos aspectos:

— *Necesidad* de la carrera en el medio, observada a través del nivel de absorción de los egresados. Para ello se realizaron encuestas a directivos de establecimientos secundarios, estatales y privados e Institutos Terciarios, y egresados de todas las carreras de la Facultad.
— *Pertinencia* de los planes de estudio, lo que se percibe por la formación de competencias necesarias para el futuro desempeño profesional. Para el análisis de cada plan se consideró su marco directriz (propósitos, avances del conocimiento científico, inclusión de todas las competencias profesionales necesarias), y las áreas curriculares que lo integran.

Las estrategias utilizadas para este análisis fueron: reuniones por Departamentos, y encuestas a docentes, egresados y alumnos.

El segundo propósito expresaba: «Posibilitar que todos los alumnos, al concluir los estudios de grado, alcanzasen los máximos niveles de logro y capacidades posibles en los diversos aspectos que configuran una formación de calidad». Esta calidad estaría definida por el desarrollo de competencias en el conocimiento disciplinar, en los sabe-

res y prácticas para el desempeño profesional, en la investigación y en lo relacionado con su adaptación a las demandas sociales y culturales.

Resultados obtenidos

Como síntesis de la evaluación de la Facultad, en relación con los propósitos mencionados, se destacó:

* Suficiencia de ámbitos de absorción de egresados de la Facultad, según una primera aproximación realizada, en carreras como Profesorados de Letras, Filosofía, Historia y Geografía.
* Adecuación relación alumno-docente (un promedio de 26 alumnos por docente).
* Adecuada cantidad de docentes en los equipos de las cátedras.
* Adecuada participación de los equipos docentes.
* Adecuada actualización de los contenidos curriculares.

- **«IDENTIFICACIÓN DE ESTUDIANTES EN RIESGO DE USO DE ESTRATEGIAS INEFICACES DE ESTUDIO»**
  UNIVERSITY OF EDINBURGH. CENTRE FOR RESEARCH ON LEARNING AND INSTRUCTION. 1996. H. TAIT & ENTWISTLE

Existía evidencia por parte de varios estudios de que el inadecuado progreso y abandono de estudios, sobre todo durante el primer año de universidad, se debía a la carencia y/o inadecuado desarrollo de estrategias y habilidades de estudio apropiadas por parte del alumno.

En recientes estudios, se preguntó al personal de universidad y colegios de educación superior, sobre programas de asesoramiento y entrenamiento en la adquisición de habilidades dirigidas al estudiante; éstos respondían que lo lógico era que los estudiantes que accedían a la universidad estuvieran ya lo suficientemente cualificados como para realizar estudios superiores.

En un reciente documento desarrollado por personal académico, se proponía que las instituciones deberían crear un plan general sobre habilidades de estudio que dejara fuera de esta responsabilidad a la institución, facultad y departamentos. La información provista en dicho plan, sería discutida por los tutores de cada asignatura, por una parte para hacerla más inteligible, pero también para ofrecer consejos sobre aspectos específicos referentes a materias o cursos concretos.

En este estudio, realizado en la Universidad de Edinburgh, se utilizaron dos sistemas de recopilación de datos: por una parte se realizaron cuestionarios e inventarios con items específicos para el propósito del proyecto, y por otra parte se utilizó un programa informatizado diseñado para que los estudiantes lo utilizaran de manera individual y completaran así el inventario, y al mismo tiempo serviría para la recopilación, en una clase completa, de datos con información para identificar aquellos alumnos con debilidades en su método de estudio y que por lo tanto necesitarían ayuda.

Para evitar que se formara la creencia de que existe una sóla forma correcta de respuesta a las cuestiones, el «studyAdvisor» aconsejó tener en cuenta que el método de estudio utilizado depende tanto del propio estudiante como del curso o área de estudio de que se trate.

Conclusión:

El asesoramiento hasta ahora ofrecido a los estudiantes, en la adquisición de habilidades de estudio, ha sido criticado y nada efectivo, sobre todo porque aparece normalmente como un anexo carente de importancia de los estudios a realizar. Para que dicho asesoramiento fuera efectivo, en primer lugar habría de ser motivador para el alumno, ser presentado como una parte esencial del curso y al que puede acudir en cualquier momento.

Sin embargo, esto sólo es posible si el personal académico está lo suficientemente familiarizado con el tipo de asesoramiento que debe dar. Por lo tanto, no hay que olvidar que instituciones, facultades y departamentos son igualmente responsables de ofrecer ayuda a los estudiantes que lo necesiten.

El programa informatizado llevado a cabo en este estudio, ayudaría al personal académico a identificar a los estudiantes que pueden necesitar ayuda, pero los estudiantes por su parte, han de acostumbrarse a utilizar el StudyAdvisor, no sólo una vez sino regularmente.

En muchos lugares, el mismo personal académico, puso de manifiesto que era necesario no sólo reforzar/reformar los programas de orientación, sino también los programas de formación, para lograr que los estudiantes desarrollasen el valor potencial que supone la mejora de sus habilidades de estudio.

- **«EXIGENCIAS DEL TRABAJO UNIVERSITARIO DESDE LA PERSPECTIVA DE ALUMNOS Y PROFESORES BIEN EVALUADOS»**
  UNIVERSIDAD DE SEVILLA. CURSO 1998/99. Informe 2000.
  Profs: Víctor Álvarez Rojo, Eduardo García Jiménez, Javier Gil Flores y Soledad Romero Rodríguez.

El estudio refleja las exigencias encontradas en una investigación realizada, durante el curso 98/99, con profesores y alumnos bien valorados de la Universidad de Sevilla y con una muestra extraída del conjunto de los docentes y discentes de esa universidad.

En este trabajo se prestó especial atención a las opiniones que profesores y estudiantes tienen sobre los factores que afectan al rendimiento discente.

La progresiva implantación de nuevos planes de estudios en la práctica totalidad de las universidades y titulaciones, ha significado en general un incremento de la dedicación y de la cuota de trabajo personal de los alumnos.

En este marco de referencia, el éxito del alumno dependerá en gran medida de su capacidad para adaptarse a las exigencias académicas universitarias y de los métodos de trabajo con los que aborda el estudio de las diferentes materias.

De acuerdo con esta idea, los expositores arriba mencionados desarrollaron el siguiente estudio, cuyos objetivos fueron identificar las principales exigencias académicas que se plantean a los alumnos universitarios.

A partir de las opiniones y experiencias personales reveladas por alumnado y profesorado, se obtuvieron una serie de dimensiones o categorías a cerca de las habilidades, capacidades o actitudes que exige el trabajo universitario.

De la información extraída se puede concluir que *estudiar en la universidad implica una serie de requerimientos básicos* (en la que profesores y alumnado coinciden):

1. Manejo de conocimientos:

— Adquirir / manipular a niveles básicos los contenidos del temario impartidos por el profesor: recuerdo y relación con otros conocimientos.
— Planificar y realizar trabajos (monográficos, sobre temas de clase, etc.)

2. Desarrollo de habilidades generales:

— Capacidad de razonar y de expresarse de palabra y por escrito.
— Capacidad para analizar críticamente la realidad.
— Capacidad para obtener información de forma autónoma: en centros de documentación o de los propios profesores.

3. Desarrollo de Actitudes:

— Vocación ante una profesión o campo profesional.
— Motivación, dedicación y responsabilidad respecto a lo que es el trabajo diario de un estudiante.
— Organización del estudio.

Conclusión:

Es necesaria una *reforma en cuanto a los programas de orientación universitaria* se refiere. Esto implica necesariamente dos primeros aspectos:

— La necesidad de desmitificar las características de una acción orientadora en la universidad.
Dicha acción es muy similar a la del nivel educativo anterior (secundaria).
— Se desprende de estos mismos datos que la *orientación universitaria no puede plantearse sin tener en cuenta a los profesores*. No será posible orientar a los estudiantes para un aprendizaje «diferente», «para acudir a la universidad con otras espectativas»...si no se orienta simultáneamente al profesorado para que enseñe y exija cosas distintas a las tradicionales.

4. CONCLUSIONES

La función esencial de la evaluación, en el aspecto que nos preocupa, es la mejora de las realidades a las que se aplica, (PÉREZ JUSTE 1997), pero puede y debe generar una compresión más profunda del objeto o campo específico al que hace referencia una intervención, lo cual significa que nos aporta «*conocimientos*» y, en esa medida contribuye al progreso científico, (DE MIGUEL 1999). Cabe por tanto considerar que uno de los objetivos fundamentales de todo proceso evaluativo —independientemente de dónde se lleve a cabo— sea mejorar la realidad donde se aplica y generar la comprensión y explicación de los problemas sociales de la misma, lo cual significa resaltar la dimensión investigadora implícita en toda evaluación (CHELIMSKY, 1997) haciendo un esfuerzo por clarificar las relaciones entre las necesidades sociales y las soluciones propuestas, entre la teoría que justifican los programas y los diseños que se formulan, entre la lógica que se utiliza para su implementación y el contexto real de la intervención.

La docencia no siempre facilita la comprensión del funcionamiento de los problemas globales que tiene el mundo, ni prepara adecuadamente y de una forma útil y práctica para la vida.

— Si uno de los objetivos fundamentales de la Universidad es preparar al estudiante para su incorporación al mercado laboral, a una sociedad en continuo proceso de transformación, a una multiculturalidad, a la internacionalización de la oferta y la demanda, etc., es la misma Universidad la primera que ha de captar todo lo que ocurre fuera para tomar las medidas necesarias de adaptación y transmisión a los estudiantes. Esto implica una *revisión continua de los Programas de Formación*, entendidos como finalidad de los procesos evaluativos.

— Al hablar de Programas de Formación, desde un punto de vista didáctico, entendemos que dicha formación no sólo se refiere a la asimilación de unos contenidos que mayoritariamente son teóricos, sino que dicha formación ha de suponer la adquisición de capacidades y competencias, lo que implica aumentar la formación en entornos naturales-reales.

— Estas capacidades y competencias deben ser *concretas*, específicas de una materia o titulación determinada, desarrolladas dentro de un buen programa de formación teórico-práctico y también, deben *globales-transversales* es decir, aquellas que preparan en aspectos más generales pero de vital importancia para el desarrollo de la carrera como pueden ser: conocimientos informáticos, movilidad intercultural, actitudes y capacidades para asumir responsabilidades y tomar decisiones, etc.

Para ello nos atrevemos a formular algunas propuestas de mejora que entre otra muchas podrían ser:

— Desarrollar estudios sobre las demandas del medio en relación con las ofertas académicas de la Universidad teniendo en cuenta las demandas del ámbito productivo para orientar los cambios.

— Ofrecer programas de formación práctica que estén actualizados y fomentar la colaboración Universidad/Empresa en todas las titulaciones.
— Fomentar los intercambios y cooperación interuniversitaria, tanto a nivel nacional como internacional, así como en cursos de grado, posgrado y prácticas profesionales en orden a la percepción directa de la homologación de la cualificación profesional de los titulados en otros países.
— Antes de iniciar la implementación de cualquier programa es conveniente estudiar las circunstancias concretas del contexto educativo en el que se va desarrollar y del contexto sociolaboral como referente potencial.
— Fomentar el desarrollo de programas de orientación adoptando modelos que se constituyan en sistemas de orientación distribuidos a lo largo y ancho de la estructura orgánica de la propia universidad en donde las funciones de orientación se corresponsabilicen entre los gabinetes de orientación universitarios, centrales y de cada centro, los departamentos, el profesorado y demás personal antes, durante y después de la realización de los estudios universitarios.
— Proporcionar a los estudiantes estrategias y habilidades generales para el empleo, su adaptación y promoción, así como el conocimiento, exploración y toma de decisiones en relación a la carrera (proceso formativo de la persona a lo largo de su vida).

Es interesante mencionar, para terminar, la adopción de modelos educativos que potencien la Transferencia del aprendizaje con vistas a conseguir una integración entre lo académico y lo profesional.

LA UNIVERSIDAD DEL FUTURO

«Cuando se encuentre en los albores de su segundo milenio la universidad, en tanto que entidad corpórea, no se parecerá mucho a lo que hasta hoy ha sido si, verdaderamente continua existiendo de forma reconocible». (Gerhard Casper, Presidente de la Universidad de Stanford. 1995.

La Universidad no puede mantenerse al margen de un entorno que se encuentra en constante evolución y transformación, esto crea la necesidad de un replanteamiento pedagógico cada vez más abierto. Pero, si como dice Barnett, 1997, la universidad está inmersa en una sociedad que requiere un continuo reajuste, lo cual parece bastante estresante, la posibilidad de establecer un equilibrio (en cuanto a programas de formación, orientación, planes de estudio, prácticas...), es cuando menos difícil.

De las aportaciones del profesor RODRÍGUEZ DIÉGUEZ, A. (2000), se deduce la necesidad de intentar desarrollar un modelo formativo impartido desde y en la universidad que promueva un tipo de competencias básicas, impregnadas de un componente crítico. Su finalidad será facilitar en los estudiantes una cierta relativización dirigida a asumir la flexibilidad suficiente, para estar abiertos a la innovación y el avance científico, consiguiendo un reajuste y adaptación constantes, ante una sociedad en permanente evolución.

450

La demanda explícita a la universidad, por parte de las instituciones sociales en las cuales se integrarán los futuros profesionales ,se concreta no tanto en un conocimiento excesivamente focalizado y especializado, sino más especialmente en el desarrollo de competencias relativas a *saber, saber ser, saber estar, saber aprender;* competencias éstas que en opinión de ECHEVERRÍA (1997) raramente se potencian en la universidad sino es a través de una perspectiva orientadora de la formación.

Esta tendencia, no sólo implica una renovación en las instituciones, sino que obliga al individuo a auto-reformarse permanentemente o mejor aún a formarse a lo largo de toda la vida. Esto nos lleva a la idea de «estudiante independiente, con un sistema de aprendizaje autónomo».

5. BIBLIOGRAFÍA

Álvarez Rojo, V., García Jiménez, E., Gil Flores, J. y Romero Rodríguez, S. (2000): *Orientación en los procesos de enseñanza-aprendizaje en la universidad*». Universidad de Sevilla. En Salmerón, H. & López Palomo, V.L. (Coords): «Orientación Educativa en las Universidades». Granada. Geu.

Apodaca, P. & Lobato, C. (1997): «*Calidad en la Universidad: Orientación y Evaluación*». Barcelona: Laertes.

Bricall, J. y Otros. (2000). «*Informe Universidad 2000*». Documento elaborado por encargo de la Conferencia de Rectores de las Universidades Españolas.

Comisión de la Comunidades Europeas. (1997): «*Cooperación Europea en Materia de Garantía de la Calidad en la Enseñanza Superior*». Bruselas. Cuadernos IRC (Revista de información sobre calidad en la universidad). «*Lineamientos generales del proceso de evaluación institucional*», «*El camino hacia el progreso*».

Consejo de Universidades (1997): «*Informe sobre la Evaluación de la Calidad de las Universidades*». Secretaría General.

Consejo de Universidades (1999). «*Guía de Evaluación para el Plan Nacional de Evaluación de la Calidad de las Universidades*». Secretaría General.

Cuadernos IRC (Revista de información sobre calidad en la universidad). «*Lineamientos generales del proceso de evaluación institucional*», «*El camino hacia el progreso*».

HEFCE (Grupo de Evaluación). (1994/95): «*Quality Assessment Report*». University of Cambridge. Architecture. Cambridge (England).

Husbands, C. (1997): «*Variations in student's evaluations of teacher's lecturing in different courses on which they lecture*». London school of Economics and Political Science. Rev. Higher Education. London (England).

Instituto Nacional de Empleo y Ministerio de Trabajo y Asuntos Sociales (1996): «*El mercado laboral de los titulados universitarios*». Servicio del Observatorio Ocupacional, cofinanciado por el Fondo Social Europeo.

Johnston, R. (1998): «*The University of the future: Boyer revisited.*». Rev. Higher education, nº 36, 253-272. Departament of Education. University of Bath, UK.

NUEC (National University Evaluation Council). (1997): «*Dividing Evaluation of Overcrowded Universities*». Evaluation in the Italian University System.

Núñez de Uberti, P. & Zidianakis, S.A. (1995): «*Evaluación de los Planes de Estudio*». Facultad de Humanidades. Ponencia dentro del «II Encuentro Interfacultades sobre la evaluación de la calidad». Secretaría General del planeamiento. Resistencia (Argentina).

Pérez Juste, R. (1997): *Evaluación de Programas.* En Salmerón Pérez, H. coord.: Evaluación Educativa. Granada. Geu:

Rodríguez Diéguez, A. (2000): «*Objetivos de la Educación para la Carrera en le ámbito Universitario*». Universidad de Sevilla. En Salmerón, H. & López Palomo, V.L. (Coords.): «Orientación Educativa en las Universidades». Granada. Geu

Rodríguez, M.L. (Coord.), Rodríguez, A., Romero, S., Padilla, T., Pereira, M. (1995): «*Educación para la carrera y diseño curricular*». Universitat de Barcelona.

Soliman, I. & Soliman, H. (1997): «*Academic workload*». University of New England. Edit. Journal Oxford. Australia.

Tait, H. & Entwistle. (1996): Centre for Research on Learning and Instruction, University of Edinburgh. Higher Education.

Tamburini, H. E. (1995): «*La visión de los docentes sobre el plan de estudio*». Facultad de Ingeniería. Universidad Nacional del Nordeste. «*II Encuentro Interfacultades sobre la evaluación de la calidad*». Secretaría General del Planeamiento. Sistema de Evaluación Permanente. (1996). Resistencia (Argentina).

Vince Thomson. Department of Mechanical Engineering. (1994): «*Value Engineering at McGILL University*». McGill University. EEUU.

452

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 463-479

NUEVAS TENDENCIAS EN LA EVALUACIÓN DE PROGRAMAS DE EDUCACIÓN MULTICULTURAL

Margarita Bartolomé Pina y Flor Cabrera Rodríguez
Universidad de Barcelona

RESUMEN

El artículo recoge una aproximación a las diversas tendencias de investigación evaluativa que son especialmente útiles en la evaluación de programas de educación multicultural. Aunque los modelos aquí presentados tienden a desarrollarse desde una perspectiva cualitativa, el acento se sitúa en su potencial para responder a las necesidades que aparecen en la evaluación de programas sociales y, en concreto, en la evaluación de programas educativos multiculturales. Se parte, por tanto, de un breve análisis de estas necesidades para, posteriormente, centrarse en el estudio de algunos posibles modelos que intentan responder más adecuadamente a esas necesidades. El artículo presenta cuatro de ellos: la evaluación participativa, la evaluación democrática y deliberativa, la evaluación colaborativa y la evaluación de programas basados en la comunidad. Los materiales han sido recogidos de la literatura sobre el tema y, especialmente, de la experiencia en evaluación de programas multiculturales de las autoras. Finalmente, desarrolla con más amplitud el estudio del «empowerment» como estrategia de evaluación en educación intercultural, presentando un ejemplo de aplicación de dicha estrategia en una evaluación de programa.

ABSTRACT

The article brings together a number of the different approaches to evaluative research which have been found to be particulary useful in the evaluation of multicultural education programs. While the models presented here have the common feature that they are developed from a qualitative perspective, emphasis is placed in their potential in responding to those needs which appear in the evaluation of social programmes, and in particular in the evaluation of multicultural education programmes. The starting point is a brief analysis of these needs which

is followed by a study of some possible models which try to offer a more adequate response to these needs.

Four such models are presented: participative evaluation, democratic and deliberative evaluation, collaborative evaluation and the evaluation of community based programmes. The materials used are drawn from a wide variety of literature related to the theme, and especially from our own experience in multicultural programme evaluation.

Finally, this article provides a fuller development of the study of «empowerment» as an evaluation strategy in intercultural education and at the same time offers an example of the application of this strategy in a programme evaluation.

DEL DIAGNÓSTICO DE LA REALIDAD EDUCATIVA MULTICULTURAL A LA EVALUACIÓN DE PROGRAMAS

Desde hace algunos años se han sucedido en nuestro país los *informes diagnósticos* de la realidad educativa tal y como se venía desarrollando en contextos multiculturales. Muchos de ellos han sido llevados a cabo con la cobertura económica del CIDE y a ellos remitimos (Varios, 1997). Otros, han ido apareciendo en la década pasada como apoyo importante a un planteamiento alternativo que incorporara adecuadamente esta dimensión en la vida de las instituciones educativas y en su planificación curricular (Véase, por ejemplo: Siguán, 1998; Carbonell (Coor); 2000). Al contrastar el desarrollo de la realidad educativa con los modelos de educación multicultural prevalentes en Europa, estos informes nos permiten apreciar y valorar indirectamente las políticas públicas que se han ido estableciendo en las diversas autonomías. Sus características metodológicas son muy distintas. Utilizan estudios de campo extensos a través de la investigación por encuesta, pero también, y cada vez con mayor frecuencia, aportan etnografías, o recogen información a través de grupos de discusión.

No hay duda de que se precisan todavía investigaciones cuidadas que recojan la evaluación de las políticas públicas en esta materia en nuestro país. De Miguel (1999: 137-161) ha desarrollado un trabajo sobre esta área, desde la perspectiva del evaluador externo, analizando algunos de los problemas que con mayor frecuencia pueden aparecer. En realidad, lo que falta en general es una clara definición de políticas públicas en el terreno de la educación multicultural, articuladas dentro de un programa más amplio de carácter interdisciplinar. EL Plà Interdepartamental D'Inmigració, aprobado por la Generalitat de Catalunya en 1995 y evaluado en 1998, podría ir en esa dirección. Dentro del mismo, podemos apreciar el papel que se le ha concedido al Programa de Educación Compensatoria, encargado en esa Autonomía, durante estos años, de llevar adelante el apoyo logístico al alumnado «en riesgo»[1] en el que se situaban las minorías culturales.

---

1 Sobre la importancia del lenguaje en los procesos evaluativos y en concreto, sobre la expresión «jóvenes en riesgo» puede leerse un excelente artículo de Madison (2000) Language in defining Social Problems and in Evaluating Social Programs. *New Directions for Evaluation*. Nº 86 pp. 17-28. Este término asocia la multiculturalidad a una situación problemática, soslayando la riqueza que supone.

451

Sin embargo, en este artículo hemos preferido centrarnos en algunos modelos actuales para la evaluación de programas en educación multicultural. En efecto, a finales de los 90 se experimentaba un cierto cansancio en relación con los estudios diagnósticos. Tanto el profesorado como las personas interesadas en el tema procedentes de otros sectores socio-formativos, advertían la necesidad de trabajar *programas concretos*, innovadores, que permitieran dar respuestas a los problemas educativos hallados en contextos multiculturales. Esta actividad trae consigo la búsqueda de *modelos de evaluación de programas* que resulten especialmente útiles en esta *etapa inicial* de su desarrollo.

Es importante situar estos modelos en el *contexto concreto* en el que se van a aplicar. Desde nuestra experiencia en evaluación de programas, hemos podido observar que los diseños de evaluación de programas, como los procesos de investigación acción, van requiriendo modificaciones en función de las personas que toman parte en ellos, de las audiencias a las que se dirigen, de las instituciones en los que se aplican, y del mismo tipo de programas. Las necesidades van cambiando y, con ellas, también la utilidad o conveniencia de utilizar uno y otro modelo de evaluación.

NUEVAS NECESIDADES SUPONEN NUEVOS MODELOS

Vamos ahora a presentar, sin ánimo de ser exhaustivas, algunas de esas necesidades, que han sido recogidas analizando la literatura sobre el tema y, muy especialmente nuestra propia experiencia.

- ***Los diferentes tipos de programas orientan la elección del modelo de evaluación de programas.*** Shwartz (2000) y Burnett, (2000) nos presentan una clasificación similar de tipos de programas en educación intercultural: Los que focalizan la información; los que focalizan cómo los escolares aprenden y los que tienen una orientación claramente social (basados en las relaciones humanas, en las habilidades de pensamiento crítico frente al racismo o en el aprendizaje cooperativo, por poner algunos ejemplos). Ahora bien, los programas de orientación social van a tener implicaciones organizativas e institucionales, junto al cambio de actitudes y prácticas, que exigen la participación más intensiva de la comunidad de referencia. Los modelos participativos y basados en la comunidad serán más adecuados para asegurar una mayor incidencia del proceso formativo.
- ***La novedad y desconocimiento del tema, por parte de los educadores y educadoras supone incluir la dimensión formativa como un elemento clave que acompaña todo el proceso.*** Esto es especialmente importante en los momentos iniciales, cuando apenas está creada la sensibilidad hacia la educación intercultural, por ejemplo, entre quienes han de aplicar el programa. Estudiar qué modelo formativo resulta más útil para conducirlo con éxito, ha constituido una de las cuestiones clave en nuestras investigaciones evaluativas (Bartolomé, coor. 1998). La evaluación colaborativa, puede favorecer esta formación.

- ***El cambio de actitudes en los propios educadores implica procesos autorreflexivos y transformadores.*** En ese sentido, las estrategias que se han utilizado para la investigación acción resultan sumamente esclarecedoras. De ahí el auge de la evaluación participativa en la actualidad.
- ***Los programas educativos multiculturales han de articularse con un proyecto social más amplio que le sirve de referencia.*** Precisamente algunos de los fracasos que hemos constatado al intentar actuar en poblaciones escolares pertenecientes a comunidades con grupos enfrentados de diferente origen socio-étnico, se deben a no haber incluido en el diseño del programa ni en su evaluación la intervención activa de la comunidad de referencia. La evaluación de programas basada en la comunidad evitaría estos fracasos.
- ***La evaluación de programas puede hacer emerger con más fuerza conflictos latentes o procesos encubiertos de fragmentación social.*** Asegurar la inclusión de todos los grupos constituye uno de los retos, junto a los procesos de reflexión y autocrítica. La evaluación democrática y deliberativa puede ser un buen camino.
- ***La diferencia cultural está vinculada en ocasiones a la desigualdad social.*** Al trabajar sobre poblaciones asimétricas en poder, influencia, prestigio y estatus social etc necesitamos desarrollar procesos que favorezcan una mayor autonomía y desarrollo de habilidades de los grupos menos favorecidos. El «empowerment» como estrategia de evaluación resulta de extraordinaria utilidad para ello.
- ***Los programas trabajan con grupos de población cuyo lenguaje y cultura pueden diferir de la del equipo evaluador.*** Desde luego se impone más que nunca el autocontrol sobre los posibles sesgos que podemos introducir las personas evaluadoras. También hacen falta enfoques más cualitativos en la evaluación de programas que favorezcan la comprensión de lo que ocurre desde el punto de vista de los y las implicados. En cuanto al propio lenguaje de la evaluación, es preciso estudiar la forma de verter los resultados de la misma en el lenguaje de las políticas públicas a fin de que puedan ser incorporados en los procesos de toma de decisiones. Hay que tener en cuenta que el lenguaje habitual de las políticas sociales es el lenguaje de los grupos dominantes. Pero el evaluador puede luchar por impedir el mantenimiento de esa situación introduciendo conceptos alternativos que favorezcan a las minorías culturales. (Madison, 2000 y Cabatoff, 2000)

Se nos abre pues, un abanico de posibilidades y sugerencias que desborda ampliamente el margen de este trabajo[2]. Vamos, sin embargo, a asomarnos a algunos de los modelos de evaluación apuntados, deteniéndonos posteriormente en una de las estra-

---

2 Algunas de las cuestiones planteadas han sido trabajadas a partir del metanálisis cualitativo realizado sobre tres estudios de casos de evaluación de un programa de educación intercultural. Véase: Bartolomé (coor) (1998) *Evaluación de un programa de educación intercultural: desarrollo de la identidad étnica en secundaria a través de la acción tutorial*. Madrid: Informe de investigación presentado al CIDE (En prensa).

456

tegias más interesantes en la evaluación de programas educativos multiculturales: «el empowerment»[3].

LA EVALUACIÓN PARTICIPATIVA

Se trata de un término cargado de múltiples significados. Cousins y Whitmore (1998) hablarán de dos grandes corrientes:

a) La evaluación participativa práctica (Practical participatory evaluation)
b) La evaluación participativa transformadora (Transformative participatory evaluation)

Sobre esta división vuelven otros autores como Brisolara (1998). Un estudio del contenido dado a estas dos corrientes permite establecer un claro paralelismo entre la 1ª y la investigación acción práctica y cooperativa (en la línea de Tikunoff y Ward, 1982) y entre la 2ª y la investigación participativa e investigación acción emancipatoria, (situadas estas dos últimas en el enfoque socio-crítico). Los autores citados y los indicadores del proceso, así lo sugieren.

De todas formas, como reconocen Cousins y Whitmore, aunque difieran en la orientación filosófica (pragmática y emancipatoria respectivamente), existen en estas dos corrientes muchos elementos que comparten y no siempre es claro situar autores y modelos en esta clasificación.

El análisis que realizan de 7 formas de evaluación participativa se lleva a cabo desde 3 ejes fundamentales:

a) Quién crea y ***controla*** la producción del conocimiento en el proceso de evaluación.
b) Cómo se realiza ***la selección*** de los «stakeholders»[4] o personas interesadas y/o afectadas por el programa.
c) Cuál es la ***profundidad y la extensión de la participación*** en el proceso evaluativo por parte de los grupos implicados.

Cada uno de ellos resulta un continuo que nos permite situar una investigación participativa concreta.

¿Cómo puede llevarse a cabo una evaluación participativa? Veamos un ejemplo por Sandín (1997) en la evaluación de un Programa de desarrollo de la identidad cultural en adolescentes. Este es el diseño propuesto

---

3 Este término aún no tiene una traducción admitida y generalizada por la comunidad científica, en castellano. De ahí que para su identificación hayamos preferido ponerlo en inglés.

4 Este término puede definirse como aquellos con intereses en el objetivo de la evaluación, es decir los sponsors del programa, gestores e implementadores, miembros de grupos de especial interés, beneficiarios del programa que tienen un espacio definido en su evaluación. Como vemos va más allá que lo que hemos definido habitualmente como «audiencia».

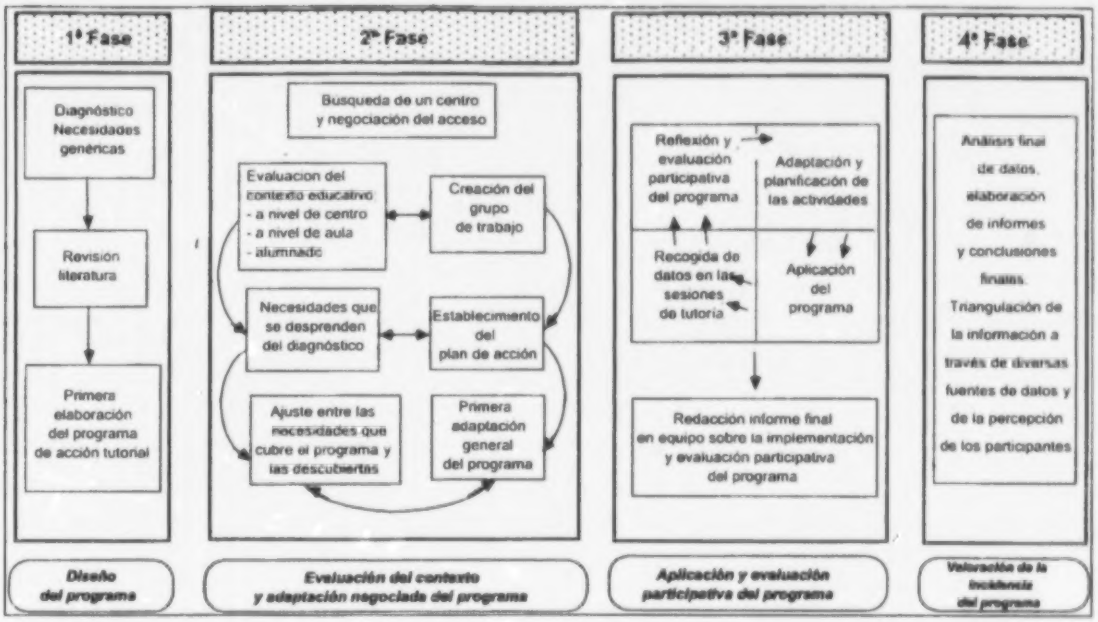


Figura 1
Diseño de la investigación: fases del proceso. (Sandín, 1997: 227)

Desde la perspectiva intercultural, queremos poner el énfasis en algunos ***elementos***, explicitados en otros trabajos (Cabrera, 2000, pp. 18-21; Sandín, 1997), que caracterizan la evaluación participativa.

- La evaluación se centra, no en la legitimación o deslegitimación sino en ***la mejora*** de grupos e instituciones. El equipo de trabajo, que ha asumido la evaluación del programa es en sí mismo un activador del cambio puesto que genera una ***actitud reflexiva y transformadora*** de sus miembros.
- Este proceso promueve ***el aprendizaje*** en la realidad sobre la que actúa.

> «*Trabajar en grupo; contrastar opiniones; discutir como enfocar diferentes actividades; valorar resultados; introducir el diario como elemento de reflexión; animarnos mutuamente en el caso de duda sobre el trabajo realizado... Son para mí aspectos positivos que me han ayudado a formarme... y esto gracias a las personas que formabamos en el grupo de trabajo (M.R. Profesora participante. Informe final.)*» (Sandín, 1997)
>
> «*Ellos*[5] *y yo hemos cambiado. Me piden más; más interacción. Ellos han ganado en sinceridad y honradez. Pero esto me ha obligado a mí a entregarme más, a no impo-*

5 Se refiere al alumnado. El grupo era muy heterogéneo respecto a su origen cultural. El profesor pertenecía a la etnia gitana.

nerles las cosas, a discutirles, a consensuarlas... La cultura del grupo ha cambiado sensiblemente porque, entre otras cosas se ha acostumbrado al debate... a hablar en público, a escuchar a los demás, a respetarse». (D. Profesor participante. Informe final. En Sandín, 1997).

- ***La participación en la evaluación*** conduce a un nivel cada vez más fuerte de ***aproximación del proyecto, de cohesión del grupo,*** una mayor ***capacidad de autorreflexión y cambio y mayor autonomía respecto a asesores externos.*** Se trata de un proceso ***paulatino y gradual.***

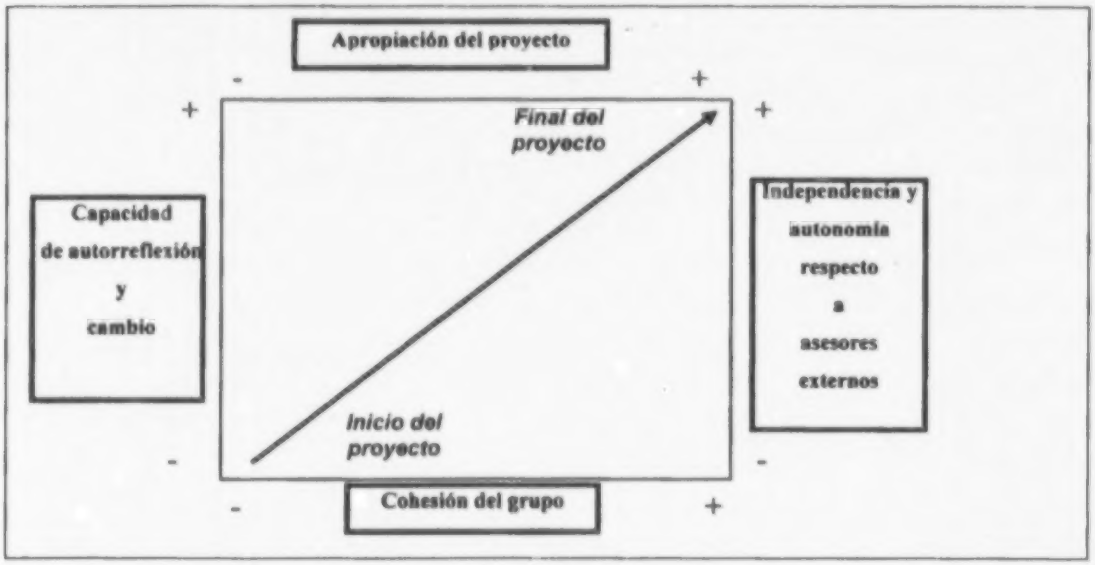


Figura 2
Evolución del equipo de trabajo a lo largo del proceso (Sandín, 1997: 388)

- ***El evaluador se convierte en un facilitador del aprendizaje que se promueve en el interior del proceso evaluativo.*** A través de estrategias de ***«empowerment»,*** las minorías culturales desarrollan habilidades de autoevaluación que les permiten no solo autoafirmarse y desarrollar una autoestima positiva sino descubrir su propia riqueza que pueden compartir con los demás:

«Me gustaba mucho porque no sé... siempre es que no sabían nada ¿No? Todos pasaban pero ahora ya... han aprendido cosas mías y me siento diferente. Yo lo veo bien porque ya que nosotros, cuando venimos a España aprendemos cosas, idiomas y todo eso... ¡Ellos también tienen derecho a aprender!» (Alumna marroquí, Centro. M. Bartolomé, coor, 1998: 562).

Dada la importancia de este rasgo lo desarrollaremos ampliamente más tarde. Existen otras formas de evaluación de programas, estan estrechamente asociadas a la evaluación participativa. Veámoslo en el cuado adjunto:

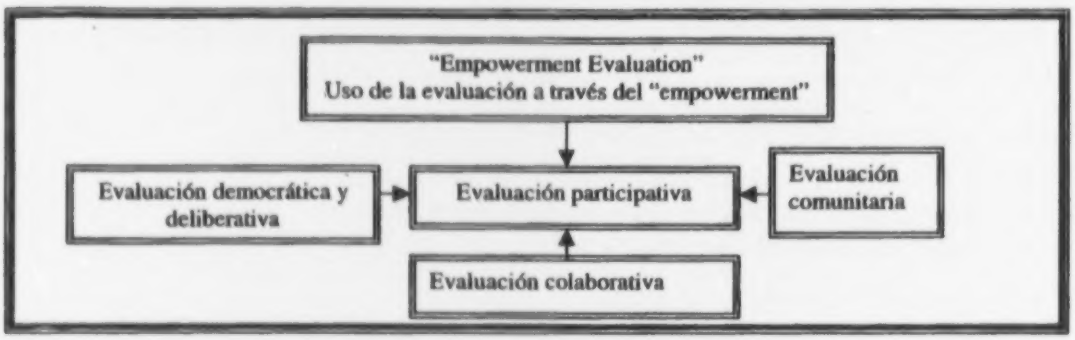


Figura 3
Tipos de evaluación asociados a la evaluación participativa

EVALUACIÓN DEMOCRÁTICA O DELIBERATIVA

Ampliamente desarrollada por McDonall (1978) para House y Kennet (2000) no constituye un modelo más sino una teoría de rango medio que sugiere que los estudios evaluativos deberían evitar un sesgo a favor de un grupo privilegiado. Se trata de legitimar el uso de la evaluación en una sociedad plural. Dado que el elemento clave en la evaluación es la representación de todos los grupos, ***la inclusión*** constituye uno de sus requisitos. Esta inclusión ha de velar porque queden representados los intereses y se asegure la participación de los que tienen menos poder en nuestra sociedad (Greene, 2000: 13). También es importante para la evaluación, su vinculación con las estructuras socio políticas más amplias, asegurando la representatividad democrática de todas las fuerzas sociales. Si el ***diálogo*** es el camino fundamental para llegar a una comunicación profunda, permitiendo dar la voz en reciprocidad, (Torres y otros, 2000: 29), la ***competencia comunicativa intercultural*** (Rodrigo, 1999), vendría asociada estrechamente a este enfoque, siendo uno de los elementos que ha de trabajarse más profundamente en contextos multiculturales.

En cuanto al tercer requisito fundamental, ***la deliberación,*** «*supone los dos elementos anteriores y permite el desarrollo del juicio crítico a partir de un proceso cognitivo fundamentado en razones, evidencias y principios válidos que faciliten la reflexión sobre las cuestiones, la valoración de sus aspectos positivos y negativos, y la búsqueda de consensos. En este proceso el evaluador, como experto, interviene aceptando que se concede autoridad pero evitando hacerse con el poder*». (Bartolomé, 2000: 26)

EVALUACIÓN COMUNITARIA

Ya indicamos en su momento la importancia de tener en cuenta la participación de la comunidad cuando tratamos de evaluar programas que incidirán sobre grupos de población que no solo difieren culturalmente, sino que pueden llegar a estar enfrentados en un conflicto de intereses sociales y/o económicos. Desde esa perspectiva, una evaluación de programa puede constituir el punto de partida de un proceso de cohesión y cambio comunitario. Ahora bien, desde la experiencia de la investigación participativa (Bartolomé y Acosta, 1992) hay que entender que la ***evaluación diagnóstica*** nos llevará una parte importante del ***proceso*** y que éste supone un largo camino para llegar a identificar los problemas fundamentales de la comunidad, así como los factores que los desencadenan, a fin de que los programas se construyan desde las ***necesidades reales***. Es importante incluir la cultura, problemas, lenguaje de las familias que forman especialmente la comunidad educativa (Bright, Epstein, Hidalgo, Siu y Swap, 1995).

Al mismo tiempo, el evaluador se convierte en «*un colaborador en los procesos que permiten el desarrollo de capacidades (habilidades, conocimientos y recursos) de la propia comunidad. Se trata de un proceso riguroso y flexible*». (Bartolomé, 2000: 27)

EVALUACIÓN COLABORATIVA

No vamos a incidir en los aspectos comunes a otros modelos ya explicitados. Queremos sencillamente subrayar dos rasgos muy interesantes:

- La importancia de plantear el diseño evaluativo a partir ***de la colaboración entre dos o más instituciones***. En nuestro caso hemos trabajado cooperativamente los centros educativos, la Universidad y el ICE como Instituto de formación del Profesorado para la evaluación de programas de educación intercultural.
- El énfasis puesto en ***la formación*** como una actividad que debe acompañar todo el proceso. Ahora bien, esta formación ha de sustentarse en la actividad autorreflexiva y crítica que desarrollan los educadores respecto al sentido y significado de sus propias acciones, sus propios prejuicios y estereotipos, su manera habitual de situarse y enjuiciar grupos y personas. A través del trabajo grupal —***seminarios formativos***[6] que acompañan los diseños de evaluación— este proceso va alcanzando un carácter ***transformador.***

En este sentido, desde la educación intercultural, hemos preferido englobar la evaluación colaborativa o cooperativa dentro de la evaluación participativa, incluyendo ***un eje más*** a los propuesto por Cousins y Whitmore: ***la dimensión formativa***.

---

6 La potencialidad de las diversas estrategias formativas puede estudiarse en la obra ya cit. Bartolomé (coor), 1998: 567 y ss así como en el informe acerca del proceso evaluativo desarrollado sobre el programa «Construïm Europa? La dimensió europea a la ESO» 1999-2000, con profesores de 4 centros de Secundaria en Barcelona. La tesis de M. Sabariego, próxima a concluirse, ofrece igualmente una aportación muy interesante en este punto, respecto al profesorado de Secundaria.

EL *EMPOWERMENT* COMO ESTRATEGIA DE EVALUACIÓN EN LA EDUCACIÓN INTERCULTURAL

La propuesta de *empowerment evaluation* que nos ofrece Fetterman (1996) constituye uno de los enfoques más prometedores del modo de hacer evaluación en contextos multiculturales, sobre todo si nos alineamos con Sleeter (1991) y concebimos el *empowerment* como un objetivo fundamental de la educación intercultural.

Aunque el término de empowerment ha sido utilizado de diferentes maneras, en el contexto de la educación asumimos el significado que le da Sleeter. Educación para el *empowerment*, según el autor, exige tomar seriamente en consideración las fortalezas, experiencias, estrategias y metas que tienen los miembros de los grupos más oprimidos, También implica ayudar a estos grupos a analizar y comprender la estructura social y a desarrollar las capacidades necesarias que les posibilite conseguir sus metas de una forma satisfactoria. Este tipo de educación trabaja a dos niveles. En el ámbito de las personas, realizar unas estrategias educativas de *empowerment* significa ayudar a las personas a aprender a utilizar la ley y los procedimientos administrativos correctos para lograr sus propias metas y objetivos y hacer valer sus intereses cuando se sienten que han sido tratados injustamente. Más que sentirse frustrados e impotentes, o dependiendo de que alguien abogue por ellos, las personas aprenden procedimientos específicos para analizar sus problemas, investigar alternativos cursos de acción, y llevar a cabo las acciones pertinentes que exige una situación justa y respetuosa. En el ámbito de grupo, esta educación significa desarrollar conocimientos, puntos de miras y habilidades para trabajar colectivamente por la justicia social. Este último *empowerment cívico* es importante en el ejemplo que posteriormente se presenta de evaluación en la medida que unos de los objetivos de ésta es contribuir a un mejor conocimiento y articulación entre asociaciones de inmigrantes y servicios municipales.

Para entender la educación multicultural como una estrategia de *empowerment* es necesario concebir el hecho educativo como un vehículo efectivo para el cambio y la emancipación social. El término multicultural en la actualidad no se asocia solamente a las respuestas educativas relacionadas con la diversidad étnico-cultural o racial, sino que amplia su campo de actuación abarcando otras formas de diversidad sobre todo las de género y de clase social. Si bien como puede observarse continúa afirmándose como una educación relacionada con grupos de personas que por alguna razón sufren algún tipo de opresión, de discriminación y/o de marginación.

La asimetría en el poder socio-político y económico que caracteriza las situaciones de estos grupos sociales en desventaja explica el énfasis en una educación intercultural que favorezca mecanismos de *empowerment* en las personas afectadas. ***Es necesario que las personas tomen conciencia crítica de la situación, de las dinámicas sociales, económicas y políticas que la generan y que reaccionen ante las mismas no como víctima ni dependiendo de otros sino potencialmente activos y con capacidad para solucionar sus propios problemas.***

Desde estas perspectivas se comprende la relevancia que en una educación intercultural tiene la propuesta innovadora de Fetterman. (1994, 1996 y 1998) de concebir

una manera de hacer evaluación que estimula el *empowerment* de las personas y organizaciones. El autor ***define empowerment evaluation como el uso de los conceptos de la evaluación, de sus técnicas y estrategias para fomentar la mejora y la autodeterminación de los participantes***. Como decíamos en otras ocasiones (Cabrera, 1998 y 2000) consiste en ***utilizar los beneficios de la metodología de la evaluación para el desarrollo de habilidades de autoevaluación que permitan a las personas, organizaciones y comunidad a mejorar por sí mismo sus actuaciones y a favorecer el cambio social necesario para que las situaciones resulten más justas y equitativas.*** Aprendiendo a ser autosuficientes para conducir sus propias evaluaciones también se es cada vez más hábil para identificar y expresar sus necesidades, establecer objetivos y expectativas, realizar un plan de acción para conseguirlos, identificar recursos, hacer decisiones racionales entre varias alternativas de acción, establecer los pasos lógicos y necesarios para conseguir los objetivos, evaluar resultados a corto y largo plazo, etc.

En este caso, la estimación del mérito o valor de un programa no es el final de la evaluación sino que, desde la perspectiva del *empowerment*, sólo es parte de un proceso en marcha para mejorar, además de las actividades, la capacidad de los participantes del programa (tanto personal responsable como usuarios, personal y organización) para ser autónomos, ejercer el control y la responsabilidad sobre lo que se hace, cómo y con qué recursos se hace y qué se consigue con ello.

Aunque Patton (1997) insiste en distinguir *empowerment evaluation* de otras formas de evaluación como son la evaluación participativa, la colaborativa, la *stakeholder involving* y la *utilización-focused*, lo cierto es que participa de muchas notas comunes entre ellas. Si bien, los distintos papeles que puede adoptar el evaluador —frente al papel clásico como experto y organizador— según las exigencias del *empowerment*, muestran que estamos ante una forma diferente de hacer evaluación; el evaluador puede ser un facilitador, un colaborador, un abogado o un formador dependiendo de las dinámicas de acción generadas por el proceso de evaluación.

Basten estas anotaciones para comprender que la capacidad de la evaluación como *empowerment* para sus participantes es crítica en el ámbito de la educación multicultural. Mediante la evaluación, muchos de estos grupos que sufren situaciones de marginación en los procesos de tomas de decisiones que les afectan pueden mejorar sus posiciones en la medida que adquieren «el poder» de su situación.

A continuación exponemos, a modo de ejemplo, las etapas del *empowerment evaluation* desarrolladas por Fetterman (1994) aplicadas a la evaluación de un programa para la inserción sociolaboral de inmigrantes del Ayuntamiento de Barcelona. Para esta presentación se hace una adaptación de la evaluación realizada por un equipo de personas del Grupo de Investigación de Educación Intercultural (GREDI) de la Universidad de Barcelona y coordinada por F. Cabrera (Cabrera y otros, 1998).

El programa *Epikouros* del Ayuntamiento de Barcelona tenía la finalidad de ayudar a personas inmigrantes no comunitarias a utilizar los servicios normalizados municipales de inserción sociolaboral y contribuir así a su integración social. Una novedad importante del programa fue el ser un proyecto cuyo éxito dependía en gran medida de la coordinación de las actuaciones de tres áreas municipales: el área

CUADRO 1

EVALUACIÓN DEL PROGRAMA EPIKOUROS DE INSERCIÓN SOCIOLABORAL DE INMIGRANTES NO COMUNITARIOS DEL AYUNTAMIENTO DE BARCELONA

ETAPAS DEL EMPOWERMENT EVALUATION	ACTIVIDADES	EMPOWERMENT COMO PROCESO	EMPOWERMENT COMO RESULTADO
PREVIO	• Formación grupo central de evaluación representativo de los distintos agentes que han intervenido en el programa: — SOLI: Servicio de orientación al inmigrante creado para el proyecto — Servicios Municipales: Instituto Municipal de Educación, Negociado de Inserción Socio laboral, Área de Promoción Económica y Ocupación (Barcelona Activa) — Asociaciones de Inmigrante: ATIME, IBN BATUTA, Asociación Dominicana de Cataluña, Asociación de Trabajadores pakistaníes, Centro Filipino, Asociación de Jóvenes y Estudiantes de Guinea Ecuatorial- — INEM	• Clarificación de las funciones de los agentes participantes en los procesos de inserción sociolaboral de los inmigrantes: — SOLI — Instituciones relacionadas con los servicios normalizados del Ayuntamiento para la inserción sociolaboral — Colectivos de inmigrantes de las minorías: magrebíes, filipinos y paquistaníes. — otros servicios: INEM, Cruz roja, Asistencia social, etc.	• Mayor conocimiento entre los distintos grupos: sus cultura y normas de funcionamiento. • Establecimiento de redes para futuras relaciones y colaboraciones: — entre los colectivos de inmigrantes entre sí — entre colectivos e instituciones — entre instituciones
PRIMERA ETAPA **Análisis del estado del Programa. Sus puntos fuertes y débiles.**	• Descripción del programa que se llevó a término: diferencias con la planificación previa realizada. — revisión de documentos sobre el programa: su fundamentación, estructura y planificación. — descripción del programa real a partir del análisis de la realidad aportada por los distintos componentes del grupo.	• Toma de conciencia de las graves consecuencia para el éxito de un programa de: — deficiencias en la planificación — falta de seguimiento del programa — no concreción y clarificación de los distintos agentes e instituciones implicadas.	• Adquisición de conocimientos sobre los elementos claves que deben considerarse en la planificación de un programa, en su implementación y seguimiento.
SEGUNDA ETAPA **Establecer los objetivos de la evaluación**	Al tratarse de una evaluación de resultados de un programa que no se pensaba repetir, las cuestiones claves que interesaban responderse que se determinaron fueron:	• Conciencia de las dificultades de concreción de resultados cuando no se disponen de unos objetivos bien formulados y cuando existen diferentes intereses que aún sin entrar en conflictos	• Adquisición de conocimientos sobre: — formulación de objetivos — transformación de objetivos en evidencias • Desarrollo de habilidades de diálogo, de resolución de conflicto, de respeto a las

	a) ¿qué puede aprenderse del proyecto que sirva para otros proyectos futuros? b) ¿en qué medida el proyecto ha sido útil para satisfacer las necesidades que motivó su creación? c) ¿cómo deberían funcionar los servicios normalizados de inserción sociolaboral del Ayuntamiento para que sean receptivos a las necesidades de las personas inmigrantes. Estas cuestiones se concretaron respecto al análisis de la eficacia del programa (utilidad del programa para los inmigrantes que lo utilizaron) y efectividad del mismo (efectos del proyecto en los servicios de inserción sociolaboral municipales y en las Asociaciones de Inmigrantes)	priorizan unos aspectos u otros. • Conciencia de las diferentes perspectivas desde la que se puede analizar la eficacia de un programa y los distintos valores sociales, humanos, económicos ... que subyacen bajo estas perspectivas.	distintas posiciones en la búsqueda de consenso entre las distintas partes. • Conocimiento entre los distintos grupos de sus máximas y mínimas aspiraciones que puede contribuir en el futuro a desarrollar una mayor aproximación en las posiciones de los distintos colectivos de inmigrantes y entre estos y la Administración Pública.
TERCERA ETAPA **Desarrollo de estrategias**	• Selección de las estrategias y procedimientos de evaluación para responder a los objetivos	• Revisión crítica de distintas estrategias de evaluación en el grupo. • Desarrollar procesos de negociación para conseguir acuerdos sobre la mejor forma de llegar a un objetivo común: consenso sobre las mejores estrategias evaluativas (técnicas de recogida de información, fuentes de información, etc.) y aceptación por los participantes y/o sus representantes. • Aprender a elaborar y analizar un cuestionario y la entrevista como técnicas de recogida de información.	• Desarrollo de la capacidad para planificar sus propias evaluaciones, utilizar con rigor algunas técnicas de evaluación y analizar información.
CUARTA ETAPA **Información del progreso e informe de resultados.**	• Elaboración de informes parciales • Elaboración de informes escritos • Presentación del informe.	• Tomar conciencia de la importancia de la devolución de la información y la necesidad de establecer retroalimentación de la información a distintas instancias como mecanismo necesario para la mejora personal y de los servicios. • Aprender a realizar informes de evaluación y su validación según las distintas audiencias. • Planificar y realizar una presentación pública de la evaluación según las distintas audiencias.	• Desarrollo de habilidades de comunicación escrita y oral. • Mayor conocimiento de los aciertos, errores y resultados del programa que le puede servir para futuras actuaciones. • Desarrollo de habilidades generales que le facilitan emprender otras acciones de una manera participativa como resultado global de todo el proceso. • Representación de sus intereses en la propuesta de un modelo de Inserción Sociolaboral para inmigrantes.

de Promoción económica y ocupación (Barcelona Activa), el Instituto Municipal de Educación y el Negociado de Inserción sociolaboral) tal como se puede observar en la figura 4.

El acceso de las personas inmigrantes al programa se realizaba a través de un Servicio de Orientación Laboral al Inmigrante (SOLI) creado específicamente para el proyecto. A este servicio le llegan personas derivadas de las distintas Asociaciones de inmigrantes que participan en el proyecto, del INEM o de los servicios sociales. El SOLI tenía la función de canalizar estas demandas y, después del diagnóstico de la situación de la persona inmigrante, configurar un itinerario formativo y llevar a cabo su seguimiento. En la mayoría de los itinerarios se comenzaba con módulos de inmersión lingüística (Área de Educación), de comprensión de la realidad de su nuevo entorno y de formación básica (cursos programados específicamente para el proyecto en el área del Negociado de Inserción sociolaboral), para posteriormente pasar a algún servicio de formación ocupacional que ordinariamente ofrece el área de Promoción económica a través de Barcelona Activa. En los casos donde la persona tenía un buen conocimiento del entorno social, se le remitía directamente a los

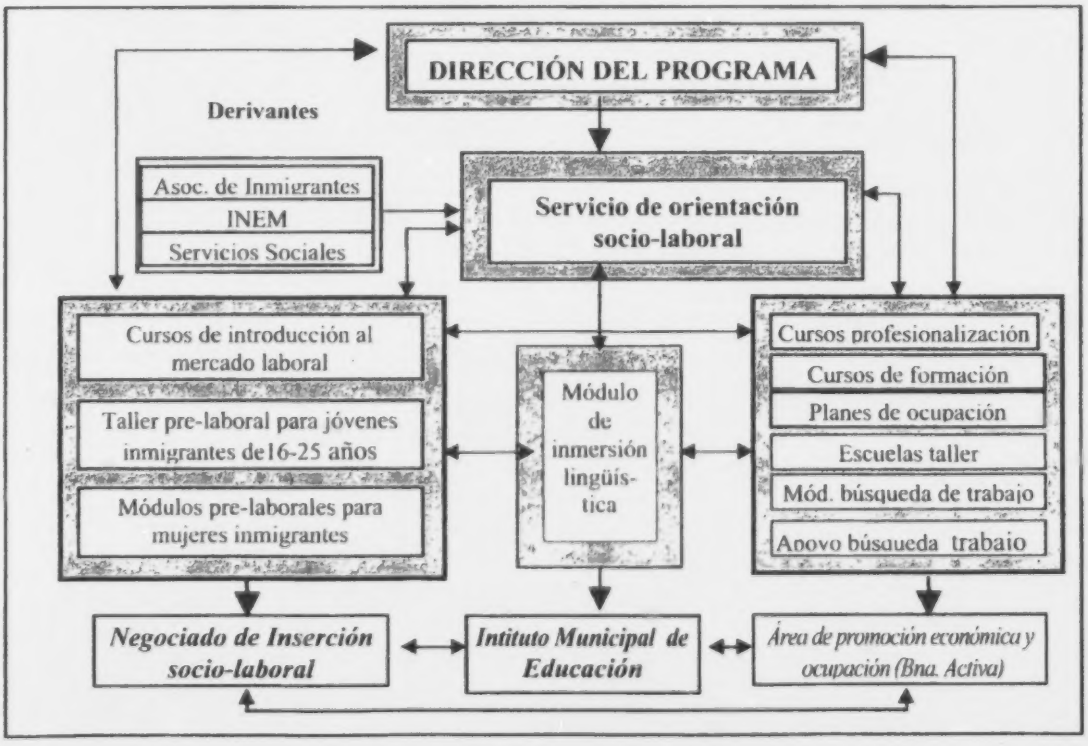


Figura 4
El programa «Epikouros» para la inserción sociolaboral de inmigrantes. (Ayuntamiento de Barcelona)

cursos de formación ocupacional u otros módulos de Barcelona Activa. El programa tuvo una duración de dos años en el transcurso de los cuales se atendió a 220 personas inmigrantes.

El ayuntamiento de Barcelona, dado que se trataba de un proyecto que había finalizado y no se continuaría con el mismo modelo de intervención, estaba interesado en una evaluación de resultados, centrada en lo que se podía aprender de la experiencia para otros proyectos futuros más que en una valoración interna del modelo de inserción sociolaboral del programa. En el Cuadro 1 se exponen las fases de la evaluación analizada desde las perspectivas del *empowerment* de los y las participantes, distinguiendo las actividades más importantes de cada etapa y una síntesis de las dinámicas del *empowerment*, como proceso y como resultado.

BIBLIOGRAFÍA

Acosta, A. y Bartolomé, M. (coord). (1992). Articulación de la educación popular con la educación formal. En *Revista de Investigación Educativa. nº* 20. 151-179.

Bartolomé, M. (Coor). (1998). *Evaluación de un programa de educación intercultural. Desarrollo de la identidad étnica en Secundaria a través de la acción tutorial.* Informe de investigación presentado al CIDE. Madrid: doc polic. 2 vol. (En prensa)

Bartolomé, M. (2000). *Hacia donde va la investigación educativa.* Lección inaugural curso 2000-2001. Facultad de Pedagogía. Barcelona: DULAC Edicions.

Brisolara, S. (1998). The History of Participatory Evaluation and Current Debates in the Field. En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 80. 25-41.

Bright, J.A.; Epstein, J.L.; Hidalgo, N.M. SIU, Sau-Fong y SWAP, S.M. (1995). Research on families, schools, and communities: a multicultural perspective. En *Handbook of research on multicultural education.* New York: Machillan. 498-524.

Burnett, G. (2000). *ERIC Clearinghouse on Urban Education: Varieties of Multicultural Education: An Introduction.* New York.

Cabatoff, K. (2000). Translating Evaluation Findings into «Policy Language». En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 86. 43-54.

Cabrera F.; Campo del, J.; Espín, J.V.; Marín M.A.; Rodríguez M. y Sandín, P. (1998). *Evaluación del proyecto Epikouros.* Informe presentado al Ayuntamiento de Barcelona. (en prensa)

Cabrera, F. (1999). Evaluación de políticas y programas sociales: de la eficacia a la participación. Ponencia en las I Jornadas de ASSCAN: *Ciudadanía, Política y Participación.* Las Palmas de Gran Canarias, junio.

Cabrera, F. (2000). *Evaluación de la formación.* Madrid: Síntesis.

Carbonell, F. (2000). *Educació i immigració. Els reptes educatius de la diversitat cultural i l'exclusió social.* Barcelona: Mediterrànea, S.L.

Coupal, F.P. y Simoneau, M. (1998). A Case Study of Participatory Evaluation in Haiti. En *New Directions for Evaluation.* Nº 80. 69-79.

Cousins, B. y Whitmore, E. (1998). Framing Participatory Evaluation. En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 80. 5-23.

De Miguel, M. (1999). Problemas teóricos y metodológicos en la evaluación externa de programas educativos multiculturales. En Martín, Mª T. y Sarrate, Mª L. *Evaluación y ámbitos Emergentes en Animación Sociocultural.* Madrid: Sanz. Pp. 137-161.

Feterman, D.M. (1997). Empowerment Evaluation and Acreditation in Higher Education. En E. Chelinsky y W. Shadish (Eds). *Evaluation for the 21st. Century. A Handbook.* London: Sage. 381-396.

Fetterman, D.M. (1996). Empowerment evaluation. An introduction to theoriy and practice. En D.M. Feterman, S.J. Kaftarian y A. Wandersman, A. (Eds): Empowerment *Evaluation. Knowledge and tools for self-assessment and accountability.* London: Sage.

Fetterman, D.M. (1994). Steps of empowerment evaluation: from California to Cape, en *Evaluation and Program Planning,* 17 (3) 305-313.

House, E.; Howe, K.R. (2000). Deliberative Democratic Evaluation. En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 85. 3-12.

Gaventa, J.; Creed, V. y Morrissey, J. (1998). Scaling Up: Participatory Monitoring and Evaluation of a Federal Empowerment Program. En *New Directions for Evaluation.* Nº 80. Pp. 81-99.

Greene, J.C. (2000). Challenges in Practicing Deliberative Democratic Evaluation. En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 85. 13-25.

Madison, A.M. (2000). Language in Defining Social Problems and in Evaluating Social Programs. En *New Directions for Evaluation.* San Francisco: Jossey-Bass Publishers. nº 86. 17-28.

McDonald, D. (1978). Evaluation and Control of Education. En D. Tawney (ed). *Curriculum Evaluation Today: Trends and Implications.* Londres: McMillan.

McTaggart, R. (1991). When Democratic Evaluation Doesn't Seem Democratic. *Evaluation Practice.* 12(1). 9-21.

Patton, M.Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice* 18 (2) 147-163.

Rodrigo, M. (1999). *Comunicación intercultural.* Barcelona: Antropos.

Sandín, M.P. (1997). *Desarrollo de la identidad étnica en adolescentes desde una perspectiva intercultural: Evaluación participativa de un programa de acción tutorial.* Universidad de Barcelona. Tesis doctoral inédita.

Torres, R.T. y Otros. (2000). Dialogue and Reflections in a Collaborative Evaluation: Stakeholder and Evaluator voices. En *New Directions for Evaluation.* Nº 85, 27-62.

Rallis, S.F. y Rossman, G.B. (2000). Dialogue for Learning: Evaluator a Critical Friend. En *New Directions for Evaluation.* Nº 86. 81-92.

Shwartz, W. (2000). *ERIC Clearinghouse on Urban Education: A Community Guide to Multicultural Education Programs.* New York. http://www.newhorizons.org/multicultura.html

Siguan, M. (1998). *La escuela y los inmigrantes.* Paidós Educador: Barcelona.

Sleeter, C.E. (ed) (1991). Empowerment through multicultural education. New York: Sage.

Smith, M.K. (1998). Empowerment evaluation: theoretical and methodolical considerations, en *Evaluation and Program Planning,* 21(3) 255-261.

Sleeter C.E. (1991). Multicultural education and empowerment. En C.S. Sleeter (Ed.). Empowerment through multicultural education. New York: University of New York.

Varios (1997). *Catorce años de investigación sobre las desigualdades en educación en España.* Madrid: CIDE.

Ward, B.A. y Tikunoff, W.J. (1982). *Research on teaching: implications for practice.* Warrenton, Virginia: US. Departament of Education. (ERIC).

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 481-491

LA EVALUACIÓN DE PROGRAMAS INFORMATIZADOS

Víctor Álvarez Rojo, Eduardo García Jiménez, Javier Gil Flores, Soledad Romero Rodríguez y Javier Rodríguez Santero

1. LA UTILIZACIÓN DE SISTEMAS INFORMATIZADOS PARA LA ORIENTACIÓN DEL APRENDIZAJE

En nuestra sociedad del conocimiento y de la información se debe contar cada vez más con las aportaciones de las Nuevas Tecnologías de la Información y la Comunicación (NTIC) como recurso para el desarrollo de la intervención orientadora. Su capacidad de almacenamiento de grandes cantidades de información, la simulación de la realidad, el establecimiento de redes de comunicación con cualquier parte del mundo y para dar un papel activo a la persona usuaria ofrece importantes posibilidades para el desarrollo de la orientación. En este sentido, Repetto y Malik (1998) señalan que el principal reto de los/as profesionales de la psicopedagogía es la adecuada utilización de estos recursos como elemento de mejora de la calidad de los procesos de orientación y para hacer a ésta más accesible a todas las personas. Tejedor et al. (1995) insisten, asimismo, en el papel igualador que debe tener la utilización de las redes de información y comunicación.

El uso de los ordenadores en el proceso de orientación puede llevarse a cabo desde diferentes niveles, los cuales responden a la propia evolución de la Informática: a) Procesamiento de información, las aplicaciones estadísticas o la corrección de pruebas psicológicas computerizadas; y, b) Sistemas más comprensivos de orientación propiamente dicha[1].

Es importante resaltar la diferencia existente entre los sistemas de información profesional asistidos por ordenador y los sistemas integrados de orientación, ya que, independientemente del uso que se les dé —se pueden utilizar conjuntamente dentro

1 Esta evolución queda claramente descrita en los trabajos de Taveira Silva y Noguer Gelma (1999) y Marco Taverner (1995).

de un cotinuum—, en sí mismos reflejan diferentes enfoques de la orientación —centrada en la información o en el individuo—.

Los *sistemas de información* enfatizan la búsqueda de ficheros que contienen información relacionada con oportunidades profesionales, educativas y sociales; estrategias de estudio, etc. Asimismo, pueden analizar información introducida por la persona usuaria y elaborar las correspondencias entre oportunidades de elección y características de los usuarios, si su finalidad es la de orientar profesionalmente; o bien entre esas características y determinadas estrategias de aprendizaje.

Los *sistemas de orientación* aprovechan las posibilidades de interactividad de la informática y en sus versiones más integradas y comprensivas se organiza la información de acuerdo con los criterios elegidos por los usuarios y las usuarias. Los sistemas integrados permiten a la persona usuaria:

— Autoevaluarse en diferentes áreas (intereses, aptitudes, valores, hábitos de estudio, estilos de aprendizaje, estrategias de aprendizaje...).
— Recoger información adaptada a sus intereses y expectativas acerca de las posibilidades profesionales, educativas y sociales.
— Aprender el proceso de toma de decisiones académicas o profesionales.
— Desarrollar habilidades, estrategias, hábitos de aprendizaje. Cuestionarse su propia concepción del aprendizaje. Aprender a aprender.

Los programas de orientación asistidos por ordenador se han visto enriquecidos en los últimos años por la aplicación de los sistemas hipermedia, los cuales, además de la combinación del ordenador y el videodisco, permiten la utilización de estructuras de información no lineales ni secuenciales, gracias al uso del hipertexto. Esta forma de organización de la información permite un mayor protagonismo a la persona usuaria y se acerca más a los procesos reales de aprendizaje. Aunque aún existen pocos programas de este tipo, se pueden encontrar ya algunos elaborados por la empresa *Career Soft* británica; en Portugal se está realizando la prueba piloto del programa *Projectos* (Taveira Silva y Noguer Gelma, 1999).

Las ventajas e inconvenientes de la utilización de los sistemas informatizados de información y orientación han sido objeto de análisis en numerosos trabajos[2]. El análisis de los inconvenientes del uso de los sistemas informáticos, más que motivar su desconsideración, lo que hace es reforzar la idea de que deben ser utilizados como un recurso más del proceso de acción orientadora, en el que seguirá siendo de especial importancia la relación personal (orientador/a-orientando/a; grupo). Por tanto, no pueden sustituir la figura del/de la orientador/a, quien deberá jugar un importante papel en sus funciones de acompañamiento afectivo, de individidualización del proceso de orientación —los sistemas se elaboran para una persona «media»— y de planifi-

2 Harris-Bowlsbey, 1991; Marco Taverner, 1995; Offer, 1997; Repetto et alt, 1994; Repetto y Malik, 1998; Taveira Silva y Noguer Gelma, 1999. Se dedica una especial atención a ello en los monográficos del año 1984 de las revistas *Journal of Counseling and Development* y *Measurement and Evaluation in counseling and Development*.

470 A

cación de la acción orientadora. Así pues, como ha señalado recientemente Boy (1999), es preciso examinar y cuidar con mucha atención la utilización del «libre acceso» a los sistemas informáticos para que realmente cumplan su función orientadora. Entre las condiciones que deben darse para un adecuado uso de los sistemas de orientación e información asistidos por ordenador, Gati (1994) destaca las que se recogen a continuación:

1. Hay que tener en cuenta que toda la información —incluida la que proviene directamente del/de la orientador/a— puede estar sujeta a una selección no intencional.
2. Necesidad de hacer explícitas tanto las aspiraciones como los compromisos.
3. Necesidad de ofrecer sólo la información relevante.
4. Necesidad de evaluar la calidad del proceso de orientación.
5. Los/as profesionales de la orientación deben contribuir a un buen uso de los sistemas informáticos a través de las siguientes acciones:

   A Crítica constructiva de los sistemas existentes.
   B Participación en la selección de los sistemas que va a utilizar el/la destinatario/a.
   C Diseño de procesos diagnósticos para determinar quién debe utilizar los sistemas, cómo, cuándo y cuál.
   D Tutorización personal de los diálogos entre los/as usuarios/as y el sistema.
   E Discusión con el/la ususario/a, integrando e interpretando la información procedente del sistema y, por tanto, haciendo ésta más significativa para el/la usuario/a.
   F Valoración crítica de las teorías en las que están basados.

Estos requisitos llevan a una condición que se debe considerar como general: la necesidad de planificar y evaluar el uso de los sistemas informáticos como parte de un proceso más amplio e integral de la orientación, donde el protagonismo lo tiene la persona destinataria y el/la orientador/a debe realizar su función de acompañamiento y de experto/a en el proceso.

2. LA EVALUACIÓN DEL PROGRAMA «ATANDO CABOS»

«MAGALLANES» es un programa de orientación dirigido al desarrollo educativo del alumnado que inicia sus estudios universitarios. Concretamente orienta a los estudiantes en relación con los métodos de trabajo y estudio propios de este nivel educativo. En su estructura es posible diferenciar dos momentos o fases diferentes: 1) una estudio autoexploratorio basado en la comparación de las respuestas del alumno a la «Escala sobre el Trabajo y el Estudio Universitarios» con el promedio de los valores alcanzados en dicha Escala por estudiantes matriculados en Centros de su misma área de conocimiento; y, tras identificar aquellas dimensiones de la Escala que requieren un

análisis más profundo, 2) Un sistema de autoanálisis informatizado que consta de once módulos (Estilos de aprendizaje, Habilidades de planificación, Metas de aprendizaje y logro, Autoconcepto, Capacidad de expresión, Habilidades para el manejo de información, Orientación y apoyo del profesorado, Metodología didáctica del profesorado, Integración en la universidad, Capacidad de adaptación y Orientación y apoyo de la institución).

En el prorama MAGALLANES se integran dos subprogramas: «ATANDO CABOS», un programa relativo a los métodos de trabajo de los alumnos que inician sus estudios en la Universidad; y, «MANEJANDO EL TIMÓN», un programa informatizado destinado a orientar la transición a la vida activa de los estudiantes que universitarios de los últimos cursos.

Lo que a continuación presentamos es un resumen del proceso seguido para evaluar el programa «ATANDO CABOS». En este resumen nos detendremos en la consideración de diferentes elementos del diseño seguido (objetivos, términos de comparación, metodología) y en algunos de los resultados obtenidos inmediatamente después de la aplicación del mismo a un grupo de alumnos universitarios.

2.1. Objetivos de la evaluación

La evaluación del programa «ATANDO CABOS» persigue una doble finalidad. De una parte, pretende determinar la valía del mismo como recurso instruccional destinado a ayudar al estudiante a analizar sus estrategias de trabajo y sus carencias en relación con las demandas académicas del estudio en la universidad. De otra parte, la evaluación aspira a proporcionar la base informativa necesaria para adoptar decisiones de mejora del propio programa. En nuestro planteamiento el primero de los objetivos mencionados queda subordinado al segundo, puesto que el interés que para nosotros tiene la emisión de juicios de valor acerca del programa es su utilidad como punto de partida para decidir qué elementos y en qué sentido habrían de ser modificados. En este sentido, cabría hablar de una evaluación orientada a la decisión (Stufflebeam y Shinkfield, 1987), que en nuestro caso llevaría a eventuales modificaciones del programa en su presentación, estructura o contenidos.

Aunque estaríamos hablando fundamentalmente de una evaluación del diseño del programa, a ella habría que añadir la evaluación de necesidades realizada con anterioridad al desarrollo del programa y una evaluación de los resultados obtenidos con el mismo. Así, el programa «ATANDO CABOS» fue desarrollado para atender a las necesidades detectadas en los alumnos universitarios en relación con sus métodos de trabajo; el diseño del programa en cuanto a estructura y contenido se ha apoyado en los resultados del estudio empírico sobre exigencias del trabajo universitario realizado en la Universidad de Sevilla (Álvarez, 2000). De igual modo, tras la aplicación del programa «ATANDO CABOS» a una muestra de estudiantes se realizó una evaluación de los resultados obtenidos en dicho programa; esta evaluación recoge las reacciones de los estudiantes después de trabajar con el programa, aunque desde luego no sobre el aprendizaje efectivo realizado por éstos (sólo tenemos sus apreciaciones) ni sobre el impacto del programa en su actividad dentro y fuera del aula.

472

472

2.2. Términos de comparación

Entendemos que la evaluación comporta en esencia una comparación entre términos, un análisis de las semejanzas o diferencias que existen entre el objeto evaluado y el referente que adoptamos para emitir un juicio valorativo. Un juicio de valor no es más que una estimación del grado en que el objeto evaluado se aproxima a las realizaciones o concreciones que tomamos como referencia y consideramos óptimas. En la evaluación del programa «ATANDO CABOS» hemos considerado al menos los siguientes términos de comparación:

1. La estructura de contenidos en función de la cual se ha diseñado el programa constituyen un primer término, que será comparado con las exigencias que plantea a los estudiantes universitarios el estudio, es decir, con los métodos, estrategias, habilidades que conducen a resultados satisfactorios en el trabajo del alumno universitario.
2. Otra comparación se establecerá entre los objetivos pretendidos por los autores del programa y las opiniones de los alumnos usuarios del mismo acerca de los logros alcanzados tras la aplicación.
3. Los estándares o normas de calidad recogidos en la literatura sobre evaluación de programas de orientación y de programas informatizados representan un término de comparación que actuará de referente para la estructura, el diseño, la aplicación o los resultados del programa «ATANDO CABOS».

2.3. Criterios de evaluación

Los criterios en función de los cuales se afirmará la valía del objeto sometido a evaluación no son otros que la proximidad existente entre los términos que se someten a comparación. Es decir, consideraremos que el programa «ATANDO CABOS» es valioso en la medida en que su diseño, aplicación y resultados se ajusten a los propósitos para los que fue creado, respondan a las necesidades de los alumnos universitarios en relación a la metodología de trabajo o sigan las normas de calidad que se esperan para este tipo de programas.

En relación a esto último, cabría esperar que un programa posea, entre otras cualidades, las que aquí enumeramos y consideramos como referentes para la evaluación. Estos criterios han sido elaborados a partir de la literatura sobre evaluación de programas de orientación (Pérez Juste, 1992; Repetto, 1987) y teniendo en cuenta las características que debe poseer el software educativo (Marquès, s.f.).

1. *Relevancia de los contenidos*. Los contenidos del programa se consideran interesantes, actuales y responden a problemas reales, demandas o necesidades formativas a los que es preciso dar respuesta.
2. *Claridad*. El programa debe contar con metas inequívocamente delimitadas y ha de resultar claro en sus contenidos y en los procesos que implica su desarrollo. Además, la claridad debe manifestarse en el lenguaje empleado, que ha de posibilitar la comprensión por parte de los usuarios.

473

3. *Coherencia*. Es decir, unidad armónica entre los elementos que lo componen, haciendo que cada parte del programa esté conectada con las restantes. Y además, consistencia entre los objetivos del programa y el diseño del mismo, entre los contenidos y las estrategias para desarrollarlos, entre las exigencias al usuario y los resultados esperados.
4. *Capacidad de motivación*. Los contenidos, las actividades propuestas por el programa y el entorno de comunicación deben ser suficientemente atractivos para despertar y mantener el interés de los usuarios hacia los contenidos.
5. *Adecuación a los usuarios*. Se espera que el programa sea adecuado para los sujetos a los que se destina, de tal forma que sus contenidos y tareas resulten accesibles para los individuos que se encuentran en un determinado nivel académico o poseen determinadas características en función de las cuales son considerados destinatarios del programa.
6. *Versatilidad*. El programa debería demostrar su funcionalidad en diferentes contextos, como ocurre por ejemplo con las diferentes áreas de enseñanza. Además, la versatilidad se ve favorecida por la posibilidad de integrarse fácilmente con otros medios didácticos o actividades previas o posteriores al propio uso del programa.
7. *Viabilidad*. El programa es viable en función del realismo de sus metas, la calidad técnica y su ajuste a las características de la institución o a los recursos disponibles.
8. *Utilidad*. Se refiere a cómo el programa atiende a las necesidades de los destinatarios del mismo. Un programa útil resultará relevante.
9. *Eficacia*. Desde el punto de vista de la eficacia, el programa debería conducir a los resultados previstos, comportando un beneficio para quienes lo han seguido y para la institución a la que pertenecen.
10. *Eficiencia*. Hacer referencia al balance coste-beneficio en la aplicación del programa, es decir, a la relación que existe entre los resultados conseguidos y la inversión de esfuerzo y recursos para lograrlos.
11. *Facilidad de uso*. Los programas deben ser suficientemente fáciles de manejar para que estén al alcance de todos los potenciales usuarios. El uso del programa se vería facilitado si posee un carácter autoexplicativo, es decir, si se da la posibilidad real de que los usuarios puedan utilizarlo inmediatamente sin tener que llevar a cabo largos y complejos procesos de instalación, lecturas de manuales o entrenamientos previos.
12. *Calidad del entorno audiovisual*. El atractivo del programa dependerá del entorno comunicativo configurado a partir de las pantallas. Se espera que tengan suficiente calidad, desde el punto de vista técnico y estético, elementos tales como que los títulos, ventanas, iconos, botones, formularios, hipervínculos, gráficos, fotografías, animaciones, vídeos, voz, música, tipografía, navegación, etc.
13. *Satisfacción del usuario*. Los destinatarios del programa deben quedar satisfechos tras la aplicación del mismo. Aunque cabe la posibilidad de que los sujetos queden satisfechos con un programa poco valioso desde el punto de vista técnico, la satisfacción puede tomarse como una evidencia más que apoya la

valía del programa, y además constituye un importante factor que facilita la aplicación del mismo.

14. *Evaluabilidad*. Esta cualidad implica la posibilidad de ser evaluable. Es decir, el programa en sí mismo, su aplicación o los resultados a los que conduce deberían ser susceptibles de evaluación. Ha de ser posible la recogida de información acerca del programa y la emisión de juicios de valor sobre el mismo que conduzcan a introducir medidas de mejora.

2.4. Metodología de evaluación

Como hemos adelantado, la evaluación se ha basado en la información recogida durante la aplicación del programa. Desde el punto de vista metodológico, la evaluación realizada sigue el diseño de un estudio descriptivo apoyado en procedimientos de encuesta y de observación. El empleo de estos métodos de investigación responde únicamente al propósito de lograr datos e informaciones en los que basar los juicios valorativos, garantizando la legitimidad de las conclusiones a las que lleguemos en la evaluación.

Los alumnos seleccionados para la aplicación del programa han sido un total de 108, extraídos de la población que cursa el primer año de sus estudios universitarios en centros de la Universidad de Sevilla, de los cuales 51 son hombres y 57 mujeres mientras que 51.9% de los mismos es usuario habitual de ordenadores y el 48.1% restante no utiliza habitualmente este tipo de herramientas. El procedimiento de selección utilizado se ajusta en esencia a lo que denominamos un muestreo por cuotas. En la configuración de las cuotas se han observado los siguientes criterios: a) la existencia de cinco áreas de enseñanza universitaria (Arte y Humanidades, Ciencias de la Salud, Ciencias Exactas y Naturales, Ciencias Sociales y Jurídicas, Ingenierías y Tecnologías); b) conseguir la presencia en cada cuota de alumnos procedentes de diferentes centros; c) estar matriculados en primer curso de cualquiera de las titulaciones incluidas en las respectivas áreas de enseñanza.

La utilización del programa «ATANDO CABOS» por parte de los alumnos que aceptaron participar en la experiencia tuvo lugar en sesiones celebradas a lo largo de la primera quincena del mes de abril de 2000 en el aula de informática de la Facultad de CC. de la Educación. Teniendo en cuenta la información aportada por los usuarios, el tiempo medio de trabajo con el programa ha sido de unas dos horas.

Durante la aplicación del programa se ha recogido información que nos proporciona la base para llevar a cabo la evaluación del mismo. Esta información ha consistido fundamentalmente en registros de observaciones realizadas por los observadores/aplicadores presentes en las sesiones de trabajo y en opiniones y valoraciones inmediatas realizadas por los alumnos tras su experiencia con el programa. La observación se ha llevado a cabo mediante la elaboración de un registro de tipo narrativo, para el que se proponía una estructura básica recogiendo los comentarios, las preguntas, las reacciones, las dificultades y otras incidencias detectadas en la aplicación.

El otro procedimiento utilizado ha sido un Cuestionario en el que se incluye, junto a cuestiones demográficas y preguntas sobre su experiencia informática previa, una

escala de valoración sobre diferentes aspectos del programa, en la que los alumnos deben responder utilizando una escala de cinco puntos —*nada, muy poco, algo, bastante, mucho*— según consideren correctas en mayor o menor grado las afirmaciones contenidas en cada uno de los 13 items propuestos. El conjunto de items recoge valoraciones sobre el contenido del programa, su utilidad, la facilidad de uso o el formato empleado.

Junto a la escala valorativa basada en afirmaciones que los alumnos deben valorar, se incluyó una escala de diferencial semántico, constituida por 7 parejas de adjetivos bipolares para los que se establecía una escala de siete puntos. Mediante este instrumento, los alumnos debían indicar su valoración sobre el interés, la utilidad, la claridad, la facilidad, el agrado que produce y la comprensibilidad del programa. Además, quedaba abierta la posibilidad de que el alumno expresase libremente sus opiniones sobre el programa mediante una indicación final para que añadiera cuantas valoraciones se estimen oportunas. Esta cuestión abierta ponía fin al cuestionario.

Los datos obtenidos por esos procedimientos fueron analizados utilizando estadísticos descriptivos y de contraste paramétrico y no paramétrico. Además se realizaron sendos análisis factoriales exploratorios para estudiar la estructura de la escala y la del diferencial semántico.

2.5. Resultados de la evaluación

Como mostraremos a continuación, las evidencias recogidas durante la aplicación del programa «ATANDO CABOS» nos llevan a hacer una valoración positiva del mismo, en aspectos relativos a su diseño y aplicación así como a los resultados que permite alcanzar. Asimismo, el programa genera en el alumnado interrogantes e inquietudes que pueden ser la base para continuar una intervención orientadora más personalizada.

2.5.1. Relativos al diseño

En lo relativo al diseño, nos centraremos en valoraciones sobre aspectos que tienen que ver con los contenidos incluidos en el programa, las estrategias utilizadas en el trabajo con el mismo, o el soporte informático utilizado. Resumidas estas valoraciones podemos concluir que:

1. *Los contenidos del «ATANDO CABOS» se corresponden con las necesidades de los alumnos en materia de métodos de trabajo*: se ha observado una contingencia entre los aspectos del diagnóstico inicial y los módulos de información desarrollados en el programa.
2. *Los contenidos del «ATANDO CABOS» son adecuados para alumnos universitarios de primer curso*: resultan claros, fácilmente comprensibles, y por tanto adaptados al nivel de los sujetos destinatarios.
3. *Las estrategias utilizadas se han adaptado satisfactoriamente al tratamiento de los contenidos informativos*: han facilitado el autoanálisis del alumno y su acceso a la información.

4. El *diseño informático facilita la interacción con el usuario*: no se dan interrupciones en la ejecución y los usuarios no se han sentido desconcertados o incapaces de seguir avanzando en el uso del programa.

También habría que destacar dentro del apartado de diseño es que el programa, dadas las características del propio soporte informático utilizado, ofrece un nivel experiencia experiencial que podríamos considerar metafórico. No obstante, la aplicación en grupos puede facilitar el análisis de casos y situaciones reales, cercanas a su cotidianidad.

2.5.2. *Respecto a la aplicación*

Las valoraciones y argumentos que mostramos en este apartado tratan de subrayar la aplicabilidad del «ATANDO CABOS», de acuerdo con la experiencia que hemos llevado a cabo. Destacaremos una serie de cualidades apreciadas en la aplicación del programa a los alumnos universitarios. Estas valoraciones quedaría resumidas en las siguientes:

1. *La aplicación del programa es viable en nuestros centros universitarios*: tanto en lo que se refiere a los requerimientos de hardware como del tiempo requerido por la aplicación
2. *El programa resulta fácil de usar por los alumnos universitarios*: El usuario puede conocer cómo avanzar y manejar el programa sin necesidad de que previamente se la haya entrenado o formado para esta tarea.
3. *Resulta agradable trabajar con el programa.* El entorno configurado por el programa «ATANDO CABOS» produce una predisposición positiva a recibir información sobre métodos de trabajo universitario.
4. *La aplicación del programa no difiere de manera importante según áreas de enseñanza*: En general, la forma en que se ha desarrollado la aplicación del programa es similar para los alumnos procedentes de carreras adscritas a las cinco grandes áreas de enseñanza universitaria.

Entendemos que otras de las virtualidades del programa es la riqueza que puede aportar si se utiliza en sesiones grupales, en las que el proceso de autoanálisis se vería enriquecido por las aportaciones de los compañeros. Asimismo, habría que destacar la importancia del acompañamiento del orientador, si es que existe, para profundizar en el análisis que facilita el programa y ofrecer apoyo afectivo.

2.5.3 *Respecto a los resultados*

La valoración de los resultados encuentra la dificultad de que buena parte de los efectos observables como consecuencia de la utilización del programa no se producen de forma inmediata. Tras la aplicación del programa puede analizarse en las reaccio-

nes de los estudiantes: la satisfacción de los usuarios o sus percepciones sobre los beneficios y utilidad del programa. En este sentido, podemos concluir que:

1. *El programa propicia un aprendizaje.* La opinión mayoritaria de los alumnos apunta a que se produce un aprendizaje como consecuencia de la utilización del programa, lo cual puede valorarse como un primer resultado positivo.
2. *El programa resulta útil para el alumno.* Además del aprendizaje que supone la experiencia de uso del «ATANDO CABOS», la segunda valoración que hacemos sobre los resultados del programa es la referente a la utilidad del mismo de cara a mejorar efectivamente la metodología de trabajo de los estudiantes, y en particular para orientar a los alumnos que inician estudios universitarios.
3. *El balance coste-beneficio del programa es positivo.* Las valoraciones que hacen los alumnos sobre este tema apuntan a que el tiempo invertido en la utilización del programa se ve compensado por los aprendizajes derivados del mismo.
4. *Los alumnos se muestran satisfechos tras la utilización del programa.* Todas las valoraciones que hemos ido recogiendo en este apartado ponen de manifiesto una opinión positiva de los usuarios, los cuales han quedado satisfechos con el contenido del «ATANDO CABOS», el modo en que funciona, la interactividad que promueve y los resultados que se derivan del trabajo con el mismo.

Los resultados obtenidos en este proceso de evaluación nos llevan a considerar la necesidad no sólo de mejorar aspectos del contenido y del diseño del programa «ATANDO CABOS» sino también, como ya hemos comentado, la de mirar con atención el modo en que se aplica dentro y fuera de las aulas. El trabajo con grupos de alumnos y el apoyo de un servicio de orientación podrían ser dos elementos claves para contribuir al éxito de propuestas como la que aquí hemos evaluado.

BIBLIOGRAFÍA

Álvarez, V. (Dir). (2000). *Propuestas del profesorado bien evaluado para potenciar el aprendizaje de los estudiantes*. Sevilla: ICE de la Universidad.

Boy, T. (1999). Quelques élements de réflexions sur le choix d'épreuves informatisées dans le domaine de l'aide à l'orientation. *L'Orientation Scolaire et Professionnelle, 28,* 253-273.

Gati, I. (1994). Computer-assited career counseling: dilemmas, problems and possible solutions. *Journal of Counseling and Development, 73,* 262-273.

Harris-Bowlsbey, J.A. (1991). The respective roles of the Counselor and the Computer in the Career Development Process. *Congreso Internacional de AIOSP,* Lisboa, septiembre.

Marco Taverner, R. (1995). La tecnología informática en el asesoramiento vocacional. En F. RIVAS (Ed.) *Manual de asesoramiento y orientación vocacional*. Madrid: Síntesis (383-420).

Marquès, P. (s.f.). [http://www.xtec.es/~pmarques/edusoft.htm]. (sin fecha de última revisión).

Offer, M. (1997). *Software for use in careers and educational guidance and careers education in the U.K.* Material policopiado.

Pérez Juste, R. (1992). *El modelo evaluativo de la reforma educativa.* Madrid: UNED.

Repetto, E. (1987). Evaluación de programas de orientación, en Álvarez Rojo, V. (Ed.). *Metodología de la orientación educativa.* Sevilla: Alfar, 247-275.

Repetto, E.; Malik, B. (1998). Nuevas tecnologías aplicadas a la orientación. En R. Bisquerra (Ed.). *Modelos de orientación e intervención psicopedagógica.* Barcelona: Praxis (363-374).

Repetto, E., Rus, V., & Puig, J. (1994). *Orientación educativa e intervención psicopedagógica.* Madrid: UNED.

Stufflebeam, D.L. y Shinkfield, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica.* Barcelona: Paidós/MEC.

Taveira, M.C.; Noguer, M. (1999). Nuevas Tecnologías. En Fundació Bosch y Gimpera (Ed.). *Material para el curso de Postgrado sobre Servicios de Orientación para la inserción profesional.* Barcelona: Les Heures.

Tejedor, F.J. et al. (1995). Las redes de información. *Revista de Investigación Educativa,* 287-314.

479

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 493-507

EL PROGRAMA DE ORIENTACIÓN «TU FUTURO PROFESIONAL» Y SU EVALUACIÓN

Elvira Repetto

Catedrática de Orientación Educativa. Dpto. MIDE. UNED

J. Antonio Gil

Profesor Titular Interino de Metodología de Investigación. Dpto. MIDE. UNED

RESUMEN

Una de las aportaciones más significativas de Super ha sido el énfasis en el hecho de que las «carreras» se desarrollan a lo largo de la vida y que las intervenciones psicopedagógicas han de capacitar a los clientes en las competencias para enfrentarse a las tareas vocacionales requeridas en las etapas del desarrollo correspondientes, especialmente en la de la adolescencia. Adoptando una perspectiva integradora afín a Super, si bien ampliada con otros enfoques recientes, la autora ha revisado los recientes avances en las teorías del consejo y la orientación para desarrollar la fundamentación del diseño del Programa Tu Futuro Profesional *—TPF—. El artículo muestra que el «Inventario del Desarrollo de la Carrera —Forma Escolar (Career Development Inventory.-School Form-CDI-SF) de Super, previamente, validado y adaptado en nuestro país en otro trabajo, puede emplearse para medir los resultados del Programa TPF y evaluar su efectividad. Este procedimiento es el que utilizan los autores en la presente investigación pre-post-test realizada con 4.991 estudiantes españoles de la Educación Secundaria Obligatoria y Post-Obligatoria.*

ABSTRACT

One of Super's most important contributions has been his emphasis on the fact that careers develop over the life span. He proposed that counselors measure career development through the construct of vocational maturity, by identifying the coping methods used in facing a vocational task at each chronological age. To complement the construct of maturation as the central process in adolescent career development he devised measures of career maturity, such as the Care-

er Development Inventory-School Form (CDI-SF). Adopting an integrated perspective akin to Super and his later called Career Maturity concept, the author has reviewed recent advances in guidance and counseling theories in order to lay the theoretical foundations for the design of her Career Development Program: Tu Futuro Profesional *—TFP— (Your Future Career). This article shows that, although the concept of career maturity has its limitations, the CDI-SF (before validated and adapted to Spanish studentes) can be used to measure the outcomes of the Program- TFP and to evaluate its effectiveness. This procedure was used in a pre-post-test research with 4991 Spanish students from the 7th to 11th grade.*

1. INTRODUCCIÓN

Uno de los puntos básicos del debate público a nivel europeo reside en la cualificación de los trabajadores en la competencias requeridas por la sociedad de las oportunidades del nuevo milenio, evitando la exclusión social de los grupos marginados. Algunos expertos en Orientación (Watts, 1999; Repetto, 1999) estiman que la clave para la nueva sociedad europea de las oportunidades estriba en la preparación de los trabajadores en las competencias profesionales a través de la aplicación de programas de orientación y educación para la carrera a lo largo de la vida. Pero hay un periodo de la vida que es fundamental para contribuir a la supresión de la exclusión social y a la cualificación y gestión de las carreras: el periodo de los 12 a los 18 años, es decir el de la Educación Secundaria, por ser la etapa preparatoria e inicial de los más jóvenes a la vida laboral.

Esta es la finalidad ultima que pretende la línea de investigación que desde 1989 se realiza en la Cátedra de Orientación Educativa de la UNED con la denominación anterior de «Modelo Comprensivo del Desarrollo Profesional» (Repetto, 1991; 1992) y más tarde con la de «Beneficios Educativos, y Sociales de la Orientación para el Desarrollo de la Carrera» (Repetto, 1999). Dentro de esta línea, además de otros proyectos de investigación y tras la revisión del estado de la cuestión, se fundamenta y elabora un programa para capacitar a los adolescentes en las competencias profesionales «(Repetto, 1990), se adapta el «Inventario del Desarrollo de la Carrera —Forma Escolar» (Career Development Inventoty-School Form-CDI-SF)— de Super (1981) y se realiza un estudio piloto (Anaya, Repetto 1998). Tras dos evaluaciones sucesivas del Programa en si y en su viabilidad realizadas por (Repetto E.; Ballesteros B.; y Málik B., 1995) y después de sus respectivas modificaciones (Repetto, 1994) se diseña la versión última del Programa «Tu Futuro Profesional» (Repetto, 1999).

El propósito del presente estudio reside en evaluar la eficacia del Programa TFP usando el CDI-SF de Super. Por ello, primero se ofrecen unas consideraciones sobre algunos conceptos teóricos y definiciones operativas del desarrollo de la carrera, tal como el de madurez de la carrera y del instrumento CDI-SF, y se presentan brevemente la fundamentación teórica, los objetivos básicos y el proceso de elaboración del Programa «Tu Futuro Profesional». En segundo lugar, como núcleo del trabajo, se describe el diseño utilizado y se analizan y discuten los resultados comparativos de

las ganancias obtenidas, con medidas de pre-post-test en las siete escalas del CDI-SF, por una muestra de estudiantes de Educación Secundaria del estado español. Los resultados de la investigación corroboran que la adaptación española de los componentes del CDI-SF pueden evaluar la eficacia del Programa «Tu Futuro Profesional», y que dicho Programa capacita a los estudiantes españoles en las actitudes y competencias específicas del desarrollo de la carrera.

2. LA ORIENTACION PARA EL DESARROLLO DE LA CARRERA Y EL PROGRAMA «TU FUTURO PROFESIONAL»

2.1. Algunas aportaciones recientes de la orientación para el desarrollo de la carrera

Es evidente que aquí cabría revisar los logros recientes que la orientación y la educación para la carrera nos ofrecen. En otras publicaciones se recogen con más detenimiento la situación actual de la teoría y la práctica de la orientación para el desarrollo de la carrera y sus últimas aportaciones tanto en Europa como en Estados Unidos (Repetto, 2000). Pero, por la evidente limitación de espacio, únicamente va a ser objeto de reflexión algunas de las concepciones teóricas y definiciones operativas que se relacionan con el trabajo que nos ocupa. De esta suerte, se van a formular unas consideraciones sobre el papel asignado a la madurez vocacional en el desarrollo de la carrera, sus nuevas re-formulaciones, y el peso asignado recientemente a las variables demográficas tales como la edad, el sexo y la cultura.

2.2. El papel central de Super en el desarrollo de la carrera

Parson (1909) fue el primero que escribió sobre la relación de las personas con las profesiones (el ajuste hombre-trabajo), cuando estuvo trabajando con los jóvenes para facilitarles su transición al mundo laboral. El método de hacer corresponder a cada profesión con una personalidad, permitió simplificar su utilización, y que se concibiese el nuevo modelo de Holland, quien en 1959 desarrolló su procedimiento de validación, más tarde actualizada (Holland, 1985).

En contraste con los modelos estructurales predominantes de aquel momento, Super (1957) centró su atención en la evolución de los comportamientos vocacionales, adaptando de Buehler (1933) el estudio sobre los estadios vitales para organizar el comportamiento vocacional a lo largo de la vida. Mientras que las elecciones vocacionales se consideraban de forma aislada, el trabajo de Super sugiere que todas esas elecciones deben tener en cuenta el pasado y el futuro de la persona. El desarrollo de la carrera lo concibe como el proceso de crecimiento y aprendizaje del sujeto, que tiene como resultado la modificación y/o el incremento del repertorio de la conductas vocacionales de una persona. Su visión particular hace referencia a que en cada estadio vital —crecimiento, exploración, establecimiento o mantenimiento y declive— se imponen ciertas tareas que deben dirigirse en función de algunos prerrequisitos del desarrollo, y que hay evidencias objetivas de que en relación a su madurez vocacional,

hay diferencias entre los individuos (Super 1983). Su concepto de prerrequisitos del desarrollo, se encuentra reflejado en el constructo de madurez vocacional como la preparación para enfrentarse con las tareas de desarrollo de cada una de las propias etapas de la vida («*readiness to cope with the developmental tasks of one`s life stage*») (Super y Jordaan, 1973, pp. 4). De acuerdo con el *Career Pattern Study* (Super 1957), los años posteriores a abandonar la Educación Secundaria Postobligatoria, están dirigidos a la exploración para lograr la madurez en la carrera estableciendo un patrón vocacional adulto.

Por una parte, el papel central de la «madurez vocacional» en el desarrollo de la carrera se está reconsiderando en la actualidad. En este sentido, conviene sólo apuntar algunas de las sugerencias aportadas por sus seguidores. Una sección especial de *The Career Development Quarterly* (1998) analiza la prospectiva de la teoría de Super. Las sugerencias de los autores no se refieren únicamente a las mejoras teóricas, sino que también muestran su utilidad para el trabajo de los orientadores prácticos. En este sentido, Herr (1997) guiado por sus estudios desde la perspectiva social y constructivista, recomienda mejorar ciertos aspectos de la teoría de Super, tales como la modificación de algunas dimensiones del autoconcepto, y manifiesta su preferencia por el término de adaptabilidad en la carrera sobre el de madurez vocacional. Savickas (1997) propone la adaptabilidad como un constructo puente que integra los diversos segmentos del desarrollo de la carrera a lo largo de la vida. En relación a la exploración Blustein (1997) defiende la perspectiva de un un «contexto enriquecido» como dimensión básica de la adaptabilidad a los roles vitales, concluyendo que el comportamiento exploratorio está constituido por una serie de destrezas relevantes para el desarrollo profesional del sujeto. Osborne, Brown, Niles y Miner (1997) en su libro titulado *Career Development Assessment and Counseling; Application of the Donald E. Super C-DAC Approach* (1997), revisan la teoría y las investigaciones que subrayan la valoración del desarrollo de carrera. En la misma línea, Vondracek y Reitzle (1998) sugieren que los modelos de proceso de la madurez vocacional deberían reemplazarse por un modelo de «momentos», que capturen la interrelación entre la persona y el ambiente.

Por otra parte, existen en la actualidad muchos proyectos de investigación que defienden la relevancia de las variables demográficas, tales como la edad, el sexo y la cultura. Por ejemplo Healy, Mitchell y Mourton (1987) han confirmado que la madurez vocacional aumenta con la edad; Luzzo (1995), afirma que las jóvenes obtienen niveles significativamente más altos de madurez vocacional que los jóvenes; y Post-Kammer (1987) aporta que los estudiantes de los cursos superiores obtienen niveles mayores en su madurez vocacional que los de cursos inferiores. Así mismo, Schmitt-Rodermund y Silbereisen (1998) afirman que se debería reconocer la importancia que las condiciones culturales e históricas tienen para el desarrollo de la carrera.

Por último, en los ensayos en honor a Donald Super, en *The Develpment Quarterly* (1994) se honró su memoria mostrándonos la valía de sus aportaciones. Como Savickas dice: «la conceptualización y medida del desarrollo de la carrera de Super ofrece un trampolín actual para una futura teoría del desarrollo de la carrera y su evaluación. (1994, p. 61.)

2.3. Estructura, objetivo general y proceso de validación del programa «Tu Futuro Profesional» —TFP—

Los programas compresivos del desarrollo de la carrera para adolescentes no están muy extendidos en España. Algunos de los programas de orientación vocacional extranjeros se han traducidos al español o adaptados a nuestro contexto, como el de Pelletier y Bujold (1984), que fue adaptado por Álvarez con el título «Tengo Que Decidirme» (1994), más tarde validado por Cruz (1997). También se han diseñado algunas intervenciones para mejorar diversos objetivos específicos de la orientación vocacional. Así contamos con la existencia de programas que ayudan a los estudiantes en el proceso de toma de decisiones, tales como el de Rodríguez Moreno «Aprender a Tomar Decisiones» (1994) o el de Delgado «Toma la iniciativa» (1995). Pero no existe un programa comprensivo e integrado en el currículum escolar, que capacite en las habilidades y competencias de los estudiantes para su desarrollo profesional.

Debido a esta carencia de programas comprensivos del desarrollo de la carrera, en 1989 se inicia en la Cátedra de Orientación Educativa la línea de investigación «Modelo Comprensivo del Desarrollo Profesional (Repetto, 1991), hoy titulada «Beneficios educativos y sociales de la Orientación para el desarrollo profesional» (Repetto, 2000). Después de revisar el estado de la cuestión, se diseña el programa «Tu Futuro Profesional» —TFP— (Repetto, 1990; 1994; 1999) para el entrenamiento de los estudiantes de 12 a 18 años, en las competencias específicas del desarrollo de la carrera.

El Programa TFP consta de los cuatro Módulos siguientes: Autoconocimiento, Toma de Decisiones, Exploración de la Carrera, y Planificación y Gestión de la Carrera. Cada módulo está organizado en Unidades dentro de las cuales se indican los objetivos de aprendizaje, las actividades, las temporalización, los pasos para su realización, los recursos necesarios, y las instrucciones generales para el tutor-aplicador del programa. A su vez, se propone una metodología activa que favorece la participación de los estudiantes, de sus familias y de los miembros de la Comunidad Educativa.

Los Módulos de Autoconocimiento, Toma de Decisiones y Exploración de la Carrera constan de tres niveles:

- Nivel I: Primer ciclo de ESO (12-14 años)
- Nivel II: Segundo ciclo de ESO (14-16 años)
- Nivel III: Bachilleratos, Ciclos Formativos de Grado Superior y Programas de Garantía Social (16-18 años).

El último módulo de Planificación y Gestión de la Carrera sólo tiene el Nivel II y III.

El objetivo general del programa reside en capacitar a los estudiantes en las competencias de auto-conocimiento y conocimiento del otro, en la toma de decisiones, y en las actitudes positivas hacia la exploración, planificación y gestión de la carrera.

La fundamentación teórica del programa, se basa principalmente en la teoría del Desarrollo de la Carrera de Super, la teoría del Aprendizaje de Krumboltz (1979, 1988

y 1993), la Constructivista de Peavy (1996), y en su última versión, también se ha considerado el enfoque multicultural y transcultural de Pederson (1985) y Sue (1998).

La perspectiva del desarrollo de Super fundamenta la estructura general del programa y la secuencia de las actividades, haciendo que éstas se correspondan con los estadios del desarrollo del adolescente. Más específicamente, este enfoque nos sirve de base para las Unidades que están dirigidas al autoconocimiento de intereses y valores vocacionales, a la toma de decisiones y algunas de la exploración, planificación y gestión de la carrera.

La teoría del aprendizaje de Krumboltz ha inspirado el diseño de las Unidades dirigidas a la mejora de las habilidades para la solución de problemas y la interacción con el entorno, incluyendo actividades de aprendizaje práctico que contribuyen a la modificación de las competencias individuales, hábitos de trabajo, creencias y valores.

La aproximación constructivista de Peavy, fundamenta las actividades en las que los estudiantes participan en pequeños grupos, considerándolos como un proyecto dinámico dedicado a la construcción de su propia personalidad y a la de los otros. De este modo, la inclusión del «Yo como Proyecto» y la co-construcción de nuestra personalidad, complementa la perspectiva de Super del autoconcepto.

Han existido tres versiones consecutivas del Programa (Repetto 1990, 1994 y 1999) durante su proceso de elaboración, debido a las sucesivas validaciones de expertos realizadas. La segunda validación de expertos se presentó en la Asociación Nacional del Desarrollo de la Carrera (NCDA), en la Conferencia en San Francisco (Repetto, Ballesteros y Málik, 1995). Los aspectos evaluados fueron la suficiencia, la relevancia y el alcance de los objetivos, el tamaño del grupo, la temporalización, los recursos, la fiabilidad, la viabilidad, y la adecuación de las actividades y la puesta en marcha del programa. Los elementos con puntuaciones más bajas se modificaron, y se añadieron otras unidades y actividades acordes con la perspectiva constructivista y adaptadas a la nueva legislación educativa y laboral española. Esta tercera versión del programa es la que se ha evaluado durante el año escolar 97-98. Los resultados de la evaluación son los que se muestran en este trabajo.

3. METODOLOGÍA

3.1. Diseño

Después de realizar las validaciones de expertos citadas y un estudio piloto, el presente trabajo pretende la evaluación de los resultados de la aplicación de TFP mediante un diseño cuasi-experimental pre-post-test de grupo control y experimental. Los grupos los forman parejas de alumnos de clases del mismo nivel educativo, donde cada clase se asigna aleatoriamente a cada uno de los grupos. El programa de orientación se aplica a los alumnos del grupo experimental durante un año académico, obteniéndose medidas independientes de cada estudiante.

El constructo de Super de la madurez vocacional («career maturity»), se evalúa con su Inventario del Desarrollo de la Carrera (Career Development Inventory School Form; CDI-SF). Se considera variable independiente la aplicación del Programa y

como variable dependiente el aprendizaje de los alumnos en las competencias para el desarrollo profesional en las que entrena el TFP evaluadas a través de las siete escalas del Inventario CDI-SF. EL programa se administra a una muestra de estudiantes españoles de ESO, Bachillerato, Ciclos Formativos de Grado Medio y Programas de Garantía Social, durante el año académico 97-98. Para estimar las ganancias del grupo experimental en comparación con el de control, se administra el CDI-SF a todos los estudiantes al comienzo y al final de dicho curso. El grupo experimental se entrena con el programa TFP, mientras que el del control no recibe ningún entrenamiento. Dado que el desarrollo de la carrera varía en función del curso o nivel en el que se encuentran los estudiantes dicha variable se incluye también en el estudio.

Las hipótesis de la investigación cabe formularlas diciendo:

1. Los estudiantes que se entrenan con el programa TFP obtienen puntuaciones significativamente más altas en las siete variables del CDI-SF, mostrando ganancias significativas en su desarrollo profesionall en comparación con los grupos que no reciben ningún tipo de entrenamiento.
2. Los estudiantes de niveles educativos superiores muestran ganancias significativas en su desarrollo de la carrera en comparación con aquellos procedentes de niveles inferiores.

3.2. Muestra

La muestra se obtiene con el apoyo de los estudiantes matriculados en el curso de Orientación e Intervención Psicopedagógica, de la Facultad de Educación de la UNED. Está formada por 4.991 estudiantes de Educación Secundaria (2.484 mujeres y 2.507 hombres) habiéndose perdido 87 casos. Los grupos se forman tomando dos clases del mismo nivel en diferentes centros educativos, asignando de manera aleatoria cada clase al grupo control (2.388) y al experimental (2.516). La edad de los participantes oscila entre los 11 y los 18 años, siendo uno de 11 años y doce de 18 años. La mayoría de los participantes pertenecen a los niveles 2º de ESO (2.826), 1º de ESO (958) y 1º de Bachillerato/Ciclos Formativos (611), reduciéndose la muestra en 3º de ESO (343), 4º de ESO (195) y 2º de Bachillerato/Ciclos Formtivos (58).

3.3. Procedimiento

Para obtener los efectos del programa TFP en las actitudes y competencias profesionales de los adolescentes, se administra a toda la muestra la adaptación española del CDI-SF de Super (Anaya y Repetto, 1998) tanto al principio como al final del curso. El CDI-SF, evalúa siete dimensiones de la madurez vocacional a través de las siete escalas siguientes: cuatro de ellas valoran las dimensiones específicas de la Planificación de la Carrera (CP —20 items—), de la Exploración de la Carrera (CE —20 items—), de la Toma de decisiones (DM —20 items—) y de la Información sobre el mundo del trabajo (WW —20 items—). Además, se incluyen dos grupos de factores de carácter conativo y cognitivo: las Actitudes para el Desarrollo de la Carrera (CDA), que combi-

na CP y CE, y el Conocimiento de las habilidades para el Desarrollo de la Carrera (CDK), que combina DM y WW. La última escala, denominada Orientación Total para la Carrera (COT), se compone de todas las descritas anteriormente aportándonos una valoración global sobre la madurez vocacional.

Las autoridades educativas, concedieron el permiso para administrar el CDI-SF a ambos grupos (control y experimental), y para aplicar el programa en los grupos experimentales. A los participantes se les asegura que los datos serán confidenciales y los resultados de las respuestas anónimos. Además se les da la opción de no participar en el estudio si no lo desean; aunque esta opción sólo le eligieron cuatro alumnos.

3.4. Análisis

Las ganancias en cada variable se miden mediante las diferencias obtenidas entre las medidas pre-test y post-test tanto en el grupo control como en el experimental. De este modo, se constituye una nueva variable: las puntuaciones medias de las ganancias. Se lleva a cabo un análisis de varianza (ANOVA) con las puntuaciones de ganancias para comprobar si existen diferencias significativas entre los estudiantes en función de las variables Grupo (Experimental/Control) y Nivel (1°-4° de ESO y Bachilleratos, Ciclos Formativos de Grado Medio y Programas de Garantía Social).

4. DISCUSIÓN DE LOS RESULTADOS

En primer lugar, se van a mostrar las puntuaciones medias de ls ganacias obtenidas por los alumnos en las siete escalas de CDI. Por una parte, las medias de los estudiantes del grupo experimental (Post-test) son mucho más altas que las puntuaciones del grupo control. Debemos prestar especial atención a las altas diferencias entre las medias de las ganancias en las variables actitudinal —CP y CE— (4.83 y 8.35), mientras que la media de las variables cognitivas (DM y WW) son muy bajas (1.08 y 0.76). Así mismo, las medias de las ganancias de las variables compuestas resultan más altas, en el grupo de variables actitudinales, que en el bloque de las cognitivas (CDA=9.65 y CDK=1.75).

. Las puntuaciones medias de la variable WW en el grupo control merece una especial atención (-0.07). Una explicación plausible para este valor negativo podría ser que la mayoría de los estudiantes en la Educación Superior tinen muy poca información sobre el mercado laboral. Esto va probablemente unido al hecho de que ningún programa de desarrollo de la carrera se aplica en los Centros Educativos y que la información sobre el mundo del trabajo sólo se les facilita durante el último año de la Educación Post-obligatoria. Las medias de las ganancias del grupo experimental no son tampoco muy altas (0.76). Sin embargo, es muy importante destacar la gran diferencia entre las medias de las variables totales COT, obteniendo tan sólo el grupo control una puntuación de 0.86, mientras que el experimental obtiene un valor de 12.24.

En segundo lugar, se analiza la media de las ganancias de los grupos de acuerdo al nivel educativo al que corresponden y se obtienen los siguientes resultados: en 4° de ESO, las ganancias son claramente mayores en el grupo experimental, por ejemplo en

la variable CP el grupo control obtiene –0.18 y el experimental 4.37, en CE el control obtiene 1.56 y el experimental 6.47; así como en la variable CDA y COT con diferencias aproximadas de 11 puntos. Estas diferencias son más claras en el primer curso de Educación Secundaria Post-obligatoria.

En tercer lugar, se realiza un análisis de varianza con las puntuaciones de las ganancias del grupo experimental en función de las diferencias obtenidas en el pre-test y post.test del CDI: WW (GC=2.78/GE=10.00), CE (GC=0.75/GE=10.32), CDA (GC=3.50/GE=21.04), y COT (GC=3.71/GE=23.28).

TABLA 1
ANÁLISIS DE VARIANZA DE LAS GANANCIAS DEL GRUPO EXPERIMENTAL EN FUNCIÓN DE LAS DIFERENCIAS PRE-TEST/POST-TEST.

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
GANANCIA DEL CP	Inter-grupos	16344.913	1	16344.913	109.709	.000
	Intra-grupos	718106.241	4820	148.985		
	Total	734451.154	4821			
GANANCIA DEL CE	Inter-grupos	52373.584	1	52373.584	74.038	.000
	Intra-grupos	3399711.3	4806	707.389		
	Total	3452084.9	4807			
GANANCIA DEL DM	Inter-grupos	1134.175	1	1134.175	102.438	.000
	Intra-grupos	53421.564	4825	11.072		
	Total	54555.739	4826			
GANANCIA DEL WW	Inter-grupos	816.529	1	816.529	66.241	.000
	Intra-grupos	59599.004	4835	12.327		
	Total	60415.533	4836			
GANANCIA DEL CDA	Inter-grupos	108400.101	1	108400.101	35.424	.000
	Intra-grupos	13718271	4483	3060.065		
	Total	13826671	4484			
GANANCIA DEL CDK	Inter-grupos	2837.879	1	2837.879	56.109	.000
	Intra-grupos	235489.928	4656	50.578		
	Total	238327.807	4657			
GANANCIA DEL COT	Inter-grupos	139344.502	1	139344.502	57.900	.000
	Intra-grupos	10353363	4302	2406.639		
	Total	10492707	4303			

Como puede observarse las F de cada una de las variables evaluadas por el CDI son muy elevadas y altamente significativas. Destacando especialmente, la de Planificación de la Carrera (F= 109.709), Toma de Decisiones (F= 102.438) y exploración de la carrera (F=74.038).

BEST COPY AVAILABLE

Por último, se obtienen los resultados del análisis de ganancias de los grupos en función del curso o nivel educativo. Los más significativos corresponden a 1º de Bachillerato que se muestran en la tabla 2.

TABLA 2
ANÁLISIS DE VARIANZA DE LAS GANANCIAS DE LOS GRUPOS EN FUNCIÓN DEL CURSO 1º DE BACHILLERATO

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
GANANCIA DEL CP	Inter-grupos	12084.107	5	2416.821	15.488	.000
	Intra-grupos	762577.852	4887	156.042		
	Total	774661.960	4892			
GANANCIA DEL CE	Inter-grupos	30482.302	5	6096.460	8.552	.000
	Intra-grupos	3452336.9	4843	712.851		
	Total	3482819.2	4848			
GANANCIA DEL DM	Inter-grupos	120.351	5	24.070	2.130	.059
	Intra-grupos	54973.444	4864	11.302		
	Total	55093.795	4869			
GANANCIA DEL WW	Inter-grupos	247.266	5	49.453	3.945	.001
	Intra-grupos	61092.852	4874	12.534		
	Total	61340.118	4879			
GANANCIA DEL CDA	Inter-grupos	45979.457	5	9195.891	2.021	.073
	Intra-grupos	20710744	4552	4549.812		
	Total	20756723	4557			
GANANCIA DEL CDK	Inter-grupos	526.284	5	105.257	1.669	.139
	Intra-grupos	298667.517	4735	63.077		
	Total	299193.801	4740			
GANANCIA DEL COT	Inter-grupos	118311.510	5	23662.302	4.443	.000
	Intra-grupos	23288991	4373	5325.633		
	Total	23407303	4378			

Merece destacarse el hecho de que sólo son significativas las ganancias del curso de 1º de Bachillerato en Planificación de la Carrera y Exploración de la carrera y en la Orientación Total para el Desarrollo de la Carrera, así como en la Información del Mundo del Trabajo.

Por tanto, de los resultados de ambos ANOVA caben deducir las siguientes conclusiones:

1. Existen diferencias significativas entre los grupos control y experimental en las puntuaciones obtenidas en todas las escalas tras la aplicación del programa TFP.

Estas diferencias son siempre a favor del grupo experimental, tal y como podemos observar en la tabla 1.

2. En función del grado, en 1º de Bachillerato existen diferencias significativas en todas las variables excepto en DM (p=0.059), CDA (p=0.073) y CDK (p=0.139), tal y como se observa en los resultados de la tabla 2.

En consecuencia, podemos afirmar que nuestra investigación confirma la hipótesis primera, en la cual se predicen las ganancias significativas en los estudiantes españoles de Educación Secundaria Obligatoria y Post-obligatoria en relación con su nivel de Autoconocimiento, el aprendizaje de la Toma de Decisiones, y la Exploración, Planificación y Gestión de la carrera. La hipótesis segunda, se demuestra, sólo respecto a 1º de Bachillerato confirmando los estudios previos excepto en las variables CP, CE, WW y COT.

5. CONCLUSIONES

Los resultados de nuestro estudio en relación a la evaluación del TFP, sugieren un potencial efecto positivo del programa para el desarrollo de la carrera en los estudiantes españoles de la Educación Secundaria Obligatoria y Post-obligatoria. Las ganancias en el post-test en relación a las puntuaciones del pre-test de los estudiantes que reciben el tratamiento con TFP, son sorprendentes y marcan un límite claro con los estudiantes que no recibieron el programa, teniendo siempre como referencia y como límite las generalizaciones en función de la muestra seleccionada. Los resultados han sido ciertamente mayores de los que los autores esperaban.

Futuros diseños de investigación, podrán comprender mejor el papel que la madurez vocacional juega en el desarrollo de la carrera de los adolescentes, para lo que se deben continuar evaluando las diferencias potenciales en relación a las variables demográficas y culturales. Necesitamos además tener presente otras variables como los factores socioeconómicos y culturales. Posteriores investigaciones deberían superar las limitaciones del presente estudio, como la falta de información sobre la realidad socioeconómica y geográfica de la muestra.

Otras limitaciones de la investigación surgen de la elaboración del programa TFP propiamente. Nuevas Unidades se deberán introducir para tener presentes las aportaciones inter-culturales y multiculturales (Pedersen, 1985; Sue, 1998). Ambas perspectivas serán útiles para capacitar a los estudiantes para llegar a ser conscientes de los efectos de las clases sociales, culturales y del género en el desarrollo de la carrera. Resulta necesario diseñar actividades para ayudar en el desarrollo de las competencias de los estudiantes (conocimientos, actitudes y habilidades), para funcionar efectivamente en una sociedad plural y ser capaces de negociar y comunicarse con personas de diversas procedencias. España tiene una gran diversidad regional de culturas y lenguas, y en los años recientes ha habido un incremento de la inmigración desde el Este de Europa, Norte de África e Iberoamérica, diversidad que se refleja en las escuelas. Con la introducción de variables demográficas en el estudio, y una nueva versión revisada del TFP, podrán clarificarse otros efectos del programa.

A pesar de las limitaciones del estudio, los resultados sugieren acciones particulares en las que el orientador pueda dirigir más adecuadamente el desarrollo de la carrera de los adolescentes. Asumiendo que la madurez vocacional se incrementa de forma natural a lo largo de los años, los orientadores deben desarrollar intervenciones sistemáticas, que se integren en la experiencia curricular, aportando a los estudiantes numerosas oportunidades para desarrollar las actitudes y las competencias para su desarrollo profesional. Evaluaciones periódicas de las competencias de los estudiantes deberían permitir a los orientadores determinar la efectividad de los programas y considerar estrategias alternativas para incrementar y mejorar el desarrollo de sus carreras.

También se necesita incrementar la conciencia de la política educativa española y de las autoridades, recordando la utilidad de estos programas de orientación para la carrera, insertarlos dentro del Currículum, y aplicarlos en todos los niveles educativos y a lo largo del ciclo vital. Encuentros internacionales de expertos en orientación y políticos, como el que tuvo lugar en Ottawa en 1999, y que se retomará de nuevo en el 2001 en Vancouver, pueden ser muy beneficiosos para la interacción entre los expertos de la política educativa y de la orientación. Hoy es un imperativo que, tanto los gobiernos como los responsables de la educación tomen conciencia de los beneficios educativos y sociales que los programas para la orientación del desarrollo profesional tienen en la sociedad del nuevo milenio y así evitar la exclusión social de los grupos marginados. Podemos concluir con las palabras del Prof. Hiebert en su publicación *Making Waves: Career Development and Public Policy*: «Muchas de las actuaciones políticas relativas a la estructura y la oferta del servicio para el desarrollo de la carrera deberían implicar cambios radicales que nos guíen exitosamente dentro de la nueva era» (1999, p. 1):

Entre otros asuntos, deberíamos trabajar en la redefinición del trabajo, teniendo en cuenta la globalización actual de nuestro mercado laboral, los nuevos roles profesionales y los servicios que la sociedad del nuevo milenio nos ofrece. Afortunadamente, el trabajo presentado en este artículo podrá contribuir a incrementar estas relaciones, y a estimular en el diseño de nuevas intervenciones psicopedagógicas que capaciten a los sujetos en su desarrollo profesional.

BIBLIOGRAFÍA

Álvarez Rojo, V. (1991a). *¡Tengo que decidirme!* (Cuaderno del alumno). Sevilla: Alfar.

Álvarez Rojo, V. (1991b). *¡Tengo que decidirme!* (Cuaderno del tutor). Sevilla: Alfar.

Anaya, D. & Repetto, E. (1994). Estudios de la madurez vocacional de los estudiantes de educación secundaria mediante el Career Development Inventory. *Educación XXI*, 1, 159-174.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Blustein, D.L. (1997a). A context-rich perspective of career exploration across the life roles. *The Career Development Quarterly*, 45, 260-274.

Blustein, D.L. (1997b). The role of work in adolescent development. *The Career Development Quarterly*, 45, 381-389.

Buehler, C. (1933). *Der menschliche Lebenslauf als psychologisches Problem*. Leipzig Germany: Hirzel.

Cruz, J.M. (1997). *Evaluación de Programas de Educación para la carrera*. Sevilla: Kronos.

Day, S. & Rounds, J. (1997). A little More Than Kin, and Less Than Kind: Basic Interests in Vocational Research and Career Counseling. *The Career Development Quarterly*, 45, 3, 207-220.

Delgado, J.A. (1995). *Toma la iniciativa*. Granada: Universidad de Granada.

Healy, C.C., Mitchell, J.M. & Mourton, D.L. (1984). Age and grade differences in CD among community college students. *The Review of Higher Education*, 10, 247-258.

HERR, E.L. (1997). Super´s Life-Span, Life-Space Approach and Its Outlook for Refinement. *The Career Development Quarterly*, 45, 238-245.

Hiebert B. & Bezanson L. (Eds), (1999). *Making Waves: Career Development and Public Policy. Papers, Proceedings and Strategies*. Ottawa: Human Resources Development Canada. Canadian Career Development Foundation. U.S. Human Promotion / Distribution by ERIC/CASS

Holland, J.L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2ª ed.). Englewood Cliffs, NJ: Prentice-Hall.

Krumboltz, J. (1979). «A social learning theory of career decision making». Revised and reprinted in A. Mitchell, G. Jonnes y J. Krumboltz (Eds). *Social Learning and career decision making* (pp. 19-49). Cranston, R.I.: Carroll Press.

Krumboltz, J. (1988). *The Career Beliefs Inventory*. Palo Alto, C.A: Consulting Psychologists Press.

Krumboltz, J. (1991). *Manual for the Career Beliefs Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Krumboltz, J. (1993). Integrating career and personal counseling. *Career Development Quarterly, 42 (2), 143-148*.

Luzzo, D.A. (1995). Gender differences in college students' career maturity and perceived barriers in career development. *Journal of Counseling & Development, 73, 319-322*.

Osborne, W.L., Brown, S., Niles, S., & Miner, C.U. (1997). *Career Development, Assessment, and Counseling: Application of the Donald E. Super C-DAC Approach*. Alexandria, VA: American Counseling Association.

Ostroff, C., & Rothausen, T.J. (1997). The moderating effect of tenure in person-environment fit: A field study in educational organizations. *Journal of Occupational and Organizational Psychology*, 70, 173-188.

Parsons, F. (1909). *Choosing a vocation*. Boston: Houghton Mifflin.

Peavy, V. (1996). A constructive Framework for Career Counseling. In T.L. Sexton & B.L. Griffin (ed). *Constructivist thinking in Counseling Practice, Research and Training*. 122-140. New York: Teachers College Press.

Pedersen, P. (1987). *Handbook of cross-cultural counseling and therapy*. Westpost: Praeger

Pelletier, D., Noiseaux, G. & Bujold, R. (1974). *Developement vocationnel et croissauce personelle: approche opèretorie*. Montreal: Mc Graw-Hill Ryerson Ltd.

Pelletier, D. & Bujold, R. (1984). *Pour un approuche èducative en orientation.* Québec: Qaëtar Morin.

Post-Kammer, P. (1987). Intrinsic and Extrinsic Work Values and Career maturation of 9th to 11th grade boys and girls. *Journal of Counseling and Development,* 65, 420-423.

Repetto, E. (1990). *Programa de entrenamiento de la carrera.* Cátedra de Orientación Educativa. Madrid: UNED

Repetto, E. (1991). Modelo comprensivo del Desarrollo Profesional. *Revista de Orientación Educativa y Vocacional,* 1, 53-67.

Repetto, E (1992). El desarrollo de la carrera a lo largo de la vida y la orientación de los recursos humanos. *Revista de Orientación Educativa y Vocacional,* 3, 17-29.

Repetto, E. (1994). El Programa de Exploración y Planificación de la Carrera. En E. Repetto, V. Rus y J. Puig. *Orientación Educativa e Intervención Psicopedagógica,* (pp. 781-798). Madrid: UNED

Repetto, E. (1994). *Programa de exploración y Planificación de la Carrera (EPCE).*Madrid: Cátedra de Orientación Educativa. UNED.

Repetto, E. (1999). Orientación y sociedad. *Revista de Investigación Educativa,* 17, 581-586.

Repetto, E. (1999). *Tu futuro profesional.* Madrid: CEPE.

Repetto, E. (2000). La orientación educativa: Situación actual de la teoría y la práctica de los orientación para el desarrollo de la carrera. En A. Medina Ruiz J.: *Orientación Educativa e Intervención Psicopedagógica,* (pp. 11-33). Jaén: Universidad de Jaén.

Repetto, E.; Ballesteros, B. & Malik, B. (1995). *Validation of Repetto's career planning education development program,* presented at the NCDA 6th National Conference - *Careers and Technology: Waves of transition.* San Francisco: NCDA.

Repetto, E., Rodríguez Moreno M.L., Romeo J. & Malik B. (1999). Career development in Spain. International Symposium 1999. In B. Hiebert & L. Bezanson, *Making Waves: Career Development and Public Policy. Papers, Proceedings and Strategies,* pp. (244-253). Ottawa: Human Resources Development Canada. Canadian Career Development Foundation. U.S. Human Promotion / Distribution by ERIC/CASS

Rodríguez Moreno, M.L. (1992). *Aprender a Tomar Decisiones.* Madrid: MEC

Savickas, M.L. (1994). Measuring career development: current status and further directions. *The Career Development Quarterly, 9, 43, 54-62.*

Savickas, M.L. (1997a). Adolescent career development in social context. *The Career Development Quarterly,* 45, 303-304.

Savickas, M.L. (1997b). Career adaptability: An integrative construct for life-span, life-space theory. *The Career Development Quarterly,* 45, 247-259.

Schmitt-Rodermund, E. & Silbereisen, R. (1998). Career Maturity Determinants: Individual Development, Social Context, and Historical Time. *The Career Development Quarterly,* 47, 16-31.

Sue, D.W., Carter, R.T., Casas, J.M., Fouad, N.A., Ivey, A.E., Jensen, M., Laframboise, T., Manese, J.T., Ponterotto, J.G., & Vázquez-Nutall, E.W. (1998). *Multicultural Counseling Competencies.* Thousand OAKA: Sage PL.

Super, D.E. (1957). *The psychology of careers.* New York: Harper & Row.

Super, D.E. & Jordaan, J.P. (1973). Career Development Theory. *British Journal of Guidance of Counseling, 1, 3-16.*

Super, D.E., Thompson, A.S., Lindeman, R.H., Jordaan, J.P. & Myers, R.A. (1979). *Career Development Inventory*. School Form. Palo Alto, CA: Consulting Psychologists Press, Inc.

Super, D.E., Thompson, A.S., Lindeman, R.H., Jordaan, J.P. & Myers, R.A. (1981). *Career Development Inventory*. Volume 1: User's Manual. Palo Alto, CA: Consulting Psychologists Press, Inc.

Vondracek, F. & Reitzle, M. (1998). The Viability of Career Maturity Theory: A Developmental-Contextual Perspective. *The Career Development Quarterly*, 47, 6-15.

WATTS, A.G. (1999). *Reshaping Career Development for the 21st Century*.Derby: University de Derby.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 509-519

EL ORIENTADOR COMO INVESTIGADOR DE LA ACCIÓN

Rafael Sanz Oro
Universidad de Granada

RESUMEN

En este artículo defendemos la concepción del orientador como un investigador de la acción. Pretendemos, mediante la evaluación de programas, obtener información que tenga impacto sobre el sistema y sus componentes (humanos y materiales) y que nos conduzca a la acción y al cambio. Es muy ventajoso para los prácticos que no sólo sean consumidores de resultados de investigación sino que también la produzcan. Para ello proponemos una serie de analogías entre las fases que corresponden a una buena práctica de la orientación y las de investigación de logros o resultados.

ABSTRACT

In this article we defend a conception of the counselor as someone who investigates action. Program evaluation is a means to obtaining information that enables us to act upon the system and its components (both material and human) and to change it. Practitioners should be not just consumers of research results but also producers. To that end we propose a number of analogies between the stages corresponding to the good practice of counseling and those corresponding to achievement research.

PRESUPUESTOS DE PARTIDA

La orientación en nuestro país comienza a adquirir rango profesional cuando se introduce en los centros educativos a partir de los años 1990 con motivo de la implantación progresiva de la reforma educativa (LOGSE). Hasta esos momentos, al igual

que en otros países donde la orientación gozaba ya de una tradición de más de 40 años de existencia, la orientación no era una verdadera profesión sino más bien una constelación de diversas profesiones: psicólogos, pedagogos, profesores, médicos, trabajadores sociales, entre otros. Precisamente en otros contextos (EEUU sobre todo), fue a partir de los años 1980 cuando el modelo de *Programas de Intervención* cambió la concepción de la orientación pasando de ser una constelación de servicios a ser concebida como un programa integrado en la institución escolar basado en los principios teóricos de la psicología del desarrollo y de las organizaciones. Esta situación nos permite, en la actualidad, concebir al orientador como un investigador de la acción o, lo que es lo mismo, defender un modelo pragmático de investigación-acción donde se integran los procesos de investigación de la orientación.

Por todo ello, consideramos que la ***evaluación de programas*** puede ser considerada como una forma específica de investigación-acción. En ese sentido podríamos definir la evaluación de programas en orientación como una investigación que busca responder a cuestiones relacionadas con la «vida real» de los prácticos ofreciendo información útil para tomar decisiones sobre el desarrollo, mejora y continuación de un programa específico en un sistema u organización escolar. De ese modo, ponemos el acento en la obtención de información que tenga impacto sobre el sistema y sus componentes (humanos y materiales) y nos conduce a la acción y al cambio.

Hay autores que reconocen que esta visión de la evaluación de programas como una forma de investigación-acción entra en conflicto con otras concepciones de lo que se entiende por evaluación de programas. Afirman que la evaluación, en ese sentido, no es investigación. Así, por ejemplo, Matuszek (1981) argumenta que la meta de la investigación es establecer la verdad, mientras que la meta de la evaluación de programas es mejorar el proceso de toma de decisión. Además sugiere que la investigación es amplia y generalizable mientras la evaluación es específica y situacional, con resultados únicos aplicables a un programa particular en un contexto particular. Nuestra conclusión, sin embargo, es que hoy en día hay un solapamiento entre ambos conceptos mucho mayor del que algunos autores quieren reconocer.

A pesar de los modelos de evaluación de programas existentes en la literatura, constituye un hecho aceptado por todos los autores que no existe una única estrategia válida y definitiva que satisfaga las necesidades de los orientadores interesados en llevar a cabo la evaluación de sus programas. Un programa educativo, en su sentido más amplio, incluye un conjunto organizado de recursos y actividades diseñado para facilitar el aprendizaje y el logro de las metas de los estudiantes (Maher y Bennett, 1984). La diversidad de tales programas necesita de la existencia de un modelo para seleccionar y realizar las evaluaciones de los mismos integrando una serie de estrategias en un enfoque más comprensivo y flexible. Desde el campo de la orientación, existen modelos comprensivos como los de Atkinson, Furlong y Janoff (1979), Lombana (1985) y Midkiff y Burke (1987) que nos describen diversas estrategias que los orientadores pueden utilizar para seleccionar los diversos tipos de evaluación que satisfagan las necesidades del sistema con sus metodologías correspondientes. No nos detenemos en estos modelos por haber sido ya ampliamente tratados en otras publicaciones (Sanz Oro, 1993, 1998). Simplemente puntualizar que en el caso del movimiento de la accountability («rendi-

ción de cuentas»), como procedimiento de evaluación, tiene esta un importante papel que jugar en la actual reforma educativa. La accountability, aunque todavía sea un concepto de difícil integración en la cultura de nuestro país, constituye un gran estímulo para fomentar un sistema de información sin el cual cualquier esfuerzo de reforma estaría operando a ciegas. Si dicho movimiento desea jugar el papel que le corresponde en el área educativa, tiene que producir un cambio en los valores culturales de los educadores de todos los niveles educativos. Los profesores tienen que percibirlo como valioso en sí mismo y utilizarlo de forma constructiva. Los administradores tienen que verlo como algo más que un mero requisito burocrático. El compromiso de la comunidad tiene que ser valorado, basado en datos y bien organizado.

Por último, la accountability puede constituir un instrumento muy útil para cualquier reestructuración educativa, si, como afirma Hansen (1993), cumple las siguientes condiciones:

1. Los sistemas de accountability tienen que ser logros basados y centrados en un conjunto de indicadores claramente definidos.
2. Las metas o estándares que fija la administración tienen que ser realistas y, por tanto, alcanzables por cada zona o centro educativo donde se aplican.
3. La cultura de los centros educativos tiene que cambiar mediante un proceso sistemático que incluye la ayuda técnica necesaria y la correspondiente formación del equipo educativo.
4. Los incentivos y las sanciones tienen que estar equilibradas de forma que las escuelas no sean humilladas o penalizadas por sus fracasos.
5. La accountability tiene que incorporar una base de datos de investigación que conecte las variables input, ouput y contextuales de manera que ofrezca información útil para decidir qué procesos instructivos funcionan mejor con grupos diferentes de estudiantes.
6. Se tiene que disponer de una serie de recursos que permita a los centros educativos que no están alcanzando las metas fijadas, puedan realizar estrategias de intervención apropiadas basadas en datos de investigación.

CONVERTIRSE EN UN INVESTIGADOR DE LA ACCIÓN

Para los prácticos puede ser muy ventajoso incorporar actividades de investigación en sus tareas diarias. Los orientadores, —por aquello de la accountability o rendición de cuentas— necesitarán introducir pronto en su actividad profesional mecanismos que sirvan para documentar o justificar que los servicios que ofrecen son eficaces. Por tanto, es muy importante que no sólo sean consumidores de resultados de investigación sino que también la produzcan. El ejemplo más claro lo constituye la evaluación de nuestros programas de intervención.

Algunos orientadores pueden ver el proceso de investigación como algo impersonal y muy remoto entre sus tareas profesionales y otros sugieren que los pasos para diseñar un proceso de investigación son semejantes a los pasos que se siguen en la práctica de las profesiones de ayuda (Monette, Sullivan y DeJong, 1990; Sanz Oro, 1998). Así lo

atestiguan también las revisiones efectuadas por Álvarez (1993), Álvarez et al. (1995), Álvarez y Hernández (1998), Borders y Drury (1992), Colás (1990), Gysbers et al. (1992), Lázaro (1983), de Miguel (1985), Repetto (1995), Sanz Oro (1999) y Worzbyt y O'Rourke (1989). En un trabajo reciente de Whinston (1996), nos propone que hay una serie de analogías entre las fases que corresponden a una buena práctica de la orientación y las de investigación de logros o resultados tal como queda reflejado en el Cuadro 1.

CUADRO 1
ANALOGÍA ENTRE LOS ESTADIOS DE LA PRÁCTICA DE LA ORIENTACIÓN Y EL PROCESO DE INVESTIGACIÓN

Estadios en un programa de orientación	Estadios en una investigación
Identificar las necesidades	Identificar la cuestión a investigar
Formular las metas a alcanzar	Formular el diseño de investigación
Diseñar las intervenciones	Determinar las medidas de logro y los métodos para asegurar la integridad del tratamiento
Llevar a cabo el programa	Recoger los datos
Evaluar el progreso	Analizar los datos
Toma de decisión	Interpretación y conclusiones

FUENTE: Whiston (1996, p. 617).

a) Identificar qué es lo que se va a investigar

Esta fase es coincidente con los inicios de todo proceso de intervención en orientación, a saber, cuáles son las necesidades de los usuarios objeto de nuestra actividad profesional. Identificar lo que vamos a investigar simplemente implica examinar lo que es pertinente en el trabajo del orientador y su contexto para formarse un determinado estado de opinión. Dentro de los presupuestos de este modelo de investigación acción, se trataría de responder a una serie de interrogantes como: ¿Es eficaz la labor del orientador? ¿Qué efectos produce? En nuestro caso, la cuestión clave que pueden preguntarse los prácticos es: ¿Mejoran nuestros usuarios como resultado de los modelos de intervención que adoptamos en el programa de orientación?*

---

* El lector puede consultar la excelente monografía editada por Walz y Bleuer (1993) donde se ofrece, a través de autores de reconocido prestigio, unas excelentes revisiones sobre resultados de investigación en una serie de áreas importantes en la profesión. En todas ellas se contemplan dos objetivos básicos. Primero buscar aquellos elementos, traducidos en resultados de investigación, que conducen a la afirmación de que la orientación y, por tanto, la tarea del orientador, es una tarea eficaz. En segundo lugar sostener que la eficacia es una entidad dinámica, nunca una entidad estática en el sentido de que «el orientador la tiene o no la tiene». Para ello es esencial su capacidad de perfeccionamiento profesional.

Igualmente es de destacar la revisión efectuada por Whiston y Sexton (1998) donde se analizan las implicaciones para la práctica profesional de 75 trabajos recientes de investigación (publicados entre 1988 y

b) Formular el diseño de investigación

Esta fase es análoga a la de formular las metas a alcanzar en el proceso de orientación. La selección del diseño de investigación a seguir y la utilización de sus correspondientes metodologías, depende de los recursos, del contexto donde trabaje el orientador y de las cuestiones que desee investigar. La finalidad de cualquier estudio es lograr unas conclusiones que estén bien fundamentadas y estas no se pueden obtener sin un diseño de investigación cuidadosamente seleccionado. A este respecto, son útiles los textos de Bartolomé (1986), Buendía, Colás y Hernández (1998), Cook y Reichardt (1986), García Ramos (1992), Gelso (1979), Heppner, Kivlighan y Wampold (1993), Marshall y Rossman (1989) Patton (1990), Pérez Juste (1992), Tejedor (1994) y Weis (1991) tanto para diseños cuantitativos como cualitativos.

Una cuestión importante a considerar en este apartado son los aspectos éticos cuando se diseña el estudio de investigación. Aquí, las organizaciones profesionales tienen mucho que decir. Un ejemplo, lo constituye el *Code of Ethics and Standards of Practice* de la American Counseling Association (ACA, 1995). En concreto, en la Sección G se analizan los aspectos relacionados con la investigación en la orientación y el asesoramiento. Uno de los dilemas éticos que plantea es el relacionado con la asignación del tratamiento al grupo de control en una investigación puramente experimental. Algunos estudios de investigación acción desarrollan su programa con este grupo después de haberlo llevado a cabo con el grupo experimental. Esto no siempre es una alternativa viable pues depende del contexto y de la naturaleza propia de la intervención a realizar. Otra alternativa que se propone consiste en cambiar el tratamiento a ambos grupos y ver cuál es más eficaz, en qué circunstancias y con qué tipo de sujetos. Por ejemplo, en un programa de toma de decisiones vocacionales un grupo utiliza el programa X y otro el programa Y esperando obtener unos resultados que nos indiquen cuál ha sido más eficaz.

c) Establecer los métodos que nos aseguren la integridad del tratamiento y las medidas de los logros

Esta fase, generalmente, debería estar incluida en la anterior, es decir, en la formulación del diseño de investigación. Sin embargo, dada la complejidad que presenta, Sexton et al. (1997) la otorgan un apartado propio.

Estos autores entienden que la integridad del tratamiento se refiere a la medida en la cual un tratamiento o una serie de intervenciones intencionadas son realmente llevadas a cabo. Un procedimiento lógico consistiría simplemente en preguntar a los orientadores qué es lo que han hecho a lo largo de todo el programa o bien que estos registren todo el proceso mediante la técnica del análisis del tiempo (Fairchild, 1986). En cualquier caso, los prácticos necesitan encontrar métodos pragmáticos para determinar si sus intervenciones están produciendo los efectos esperados.

1995) y cuyos resultados giran alrededor de los cuatro componentes del modelo de Gysbers y Henderson (1994). Por último, merece una especial consideración en nuestro país las aportaciones notables de los trabajos publicados en el nº monográfico de la *Revista de Investigación Educativa* (RIE) dedicado a la orientación y coordinado por el profesor Rodríguez Espinar (1998).

499

Estos efectos no se pueden obtener sin la existencia de unas medidas sólidas que nos permitan comprobar el cambio producido. Esto, sin embargo, no siempre es tan fácil. De hecho, muchos resultados varían dependiendo si la eficacia es valorada por los usuarios, orientadores o evaluadores u observadores externos. Lambert, Ogles y Masters (1992) establecieron un esquema conceptual para categorizar medidas de resultados que facilita a los prácticos e investigadores que puedan seleccionar instrumentos que generen datos desde una amplia variedad de dominios y puntos de vista (Cuadro 2). La primera dimensión es el contenido, con las categorías de rol intrapersonal (afecto, conducta y cognición), interpersonal y social. La segunda dimensión es la fuente o fuentes de información de resultados e incluye autoinforme del usuario, evaluación del orientador, observador capacitado y otras fuentes relevantes o institucionales. La tercera dimensión viene definida por los métodos o tecnologías utilizadas para la recogida de datos que pueden ser categorizados en evaluación, descripción, observación y status. La última categoría es la orientación temporal e implica si el instrumento intenta medir una característica estable parecida a un rasgo o una característica inestable parecida a un estado. El hecho de seleccionar muchas medidas de resultado es para obtener una información comprensiva. El esquema organizacional que representa el Cuadro 2 puede ayudar a los investigadores en la selección de diversos métodos para medir los resultados.

CUADRO 2
ESQUEMA DE CLASIFICACIÓN DE MEDIDAS DE RESULTADO

Contenido	Fuente	Tecnología	Orientación temporal
1. Intrapersonal Afecto Conducta Cognición	1. Autoinforme	1. Evaluación	1. Rasgo
2. Interpersonal	2. Evaluación del orientador	2. Descripción	2. Estado
3. Rol social	3. Observador capacitado	3. Observación	
	4. Relevante	4. Status	
	5. Institucional		

FUENTE: Whiston (1996, p. 618).

d) Recogida de datos

En esta fase la cooperación entre los orientadores y los clientes o usuarios es muy importante. Un proyecto puede mejorar notablemente si cuenta con la participación de los orientadores desde los inicios del proceso de investigación. Si estos orientadores

colaboran en los estadios de planificación será más problable que continuen despues con la tarea laboriosa de recogida de datos. Igualmente los usuarios/clientes se sentirán más implicados si comprenden la finalidad que se persigue en el proyecto de investigación. Por ejemplo, Johnson y Whitfield (1991) consideran que hay tres procedimientos de recogida de datos. A través de registros escritos, orales y mediante la observación de conductas.

1º) *Registros escritos*. Que pueden ser tan simples como, por ejemplo, enumerar los cinco pasos en la toma de una decisión vocacional o identificar seis recursos de ayuda que pueda disponer un estudiante con problemas personales. Las respuestas pueden ser validadas por el profesor a través de un examen o encuesta en la clase.

2º) *Registros orales*. A través de discusiones en clase, pequeños grupos o individualmente. Por ejemplo, los orientadores o profesores podrían demandar a los estudiantes que describan los pasos en la búsqueda de un trabajo, cómo solicitar una beca, cómo acceder a una ayuda para un determinado problema personal o de aprendizaje. Las respuestas obtenidas validan la adquisición del conocimiento.

3º) *Observación de conductas*. Estas observaciones pueden ser obtenidas mediante el conocimiento, actitudes y habilidades conseguidas por el alumno a través de las actividades diseñadas en el programa de orientación. Pueden ser hechas por los orientadores, profesores, padres, etc... Como ejemplos podemos citar la utilización de recursos de carácter vocacional, reunirse con nuevos grupos sociales, ser más activo en las discusiones de clase, implicarse en actividades comunitarias, etc...

No obstante, hay una serie de consideraciones éticas que hay que tener en cuenta. Nos referimos al consentimiento de los participantes y a la confidencialidad de los datos. Así lo recoge, por ejemplo, la Sección G.2 del *Code of Ethics and Standards of Practice* (ACA, 1995) donde se contempla que «... la información obtenida de los participantes durante el proceso de investigación es confidencial» (p. 38).

e) Análisis de datos

Algunos prácticos, en esta fase, se asustan y se niegan a participar en ella porque ven la investigación sujeta a análisis estadísticos excesivamente complejos. Hoy en día, la existencia de paquetes estadísticos por ordenador hacen esta labor menos tediosa. Además, si el diseño es cualitativo se utilizan métodos descriptivos lo que produce un menor rechazo. En cualquier caso, la consulta a formadores, investigadores o estadísticos facilita considerablemente la labor.

f) Interpretación y conclusiones

Esta fase es, de hecho, la más gratificante ya que se trata de comprobar si, en función de los datos de que disponemos, nuestro programa de intervención ha producido los

efectos que deseabamos conseguir. Sin embargo, la interpretación de resultados es una tarea compleja que va más allá de encontrar diferencias estadísticamente significativas en los mismos. La interpretación estadística tradicional está siendo ultimamente criticada porque no garantiza que los resultados tengan un significado práctico. Algunos investigadores están proponiendo nuevos métodos para interpretar los datos que reflejan más fielmente esa significación práctica. La interpretación de resultados de un proceso de investigación acción está fundamentada en las características propias de dicha investigación y en el significado que adquieren los resultados para los usuarios de un determinado contexto. Los resultados obtenidos necesitan ser cuidadosa y objetivamente examinados para que las conclusiones sean válidas. No olvidemos que nuestra intención final es la de evaluar aquello que ofrecemos a nuestros usuarios y que, por tanto, los resultados que obtenemos constituyen claros indicadores de la calidad de los servicios.

Así, por ejemplo, Johnson y Whitfield (1991) contemplan la existencia de tres niveles de resultados producto de un programa de orientación: resultados *inmediatos* (competencias) de un proceso específico, resultados *intermedios o terminales* (metas) cuando se finaliza un ciclo del sistema educativo y resultados a *largo plazo* que se refiere a estudios de seguimiento de un alumno ya fuera del sistema. A continuación, presentamos un ejemplo de resultados *inmediatos* (Cuadro 3) donde se incluye los resultados a alcanzar (***competencias***), ***qué*** es lo que tiene que ser aprendido, ***quién*** es el responsable, ***cuándo*** se llevará a cabo, el ***criterio*** establecido y ***cómo*** esa conducta será validada.

CUADRO 3
RESULTADOS INMEDIATOS / COMPETENCIAS

Competencia	Qué	Quién	Cuándo	Criterio	Cómo
Los estudiantes de 4º de E.S.O. demostrarán habilidades para usar fuentes de información vocacional.	Utilizar un mínimo de cuatro recursos para encontrar los siguientes elementos para la elección de una carrera: requisitos educativos, físicos, existencia local, remuneración, posibilidades profesionales.	Orientador	Al final de una unidad de planificación vocacional en una clase de lengua de 4º de E.S.O. o en la hora semanal de tutoría	· Usar cuatro recursos · Identidicar al menos una elección vocacional. · Señalar cuatro de los cinco requisitos señalados anteriormente.	El estudiante elabora un informe vocacional donde enumera los criterios y referencias vocacionales. Es validado como correcto por el orientador.
Los estudiantes de 4º de E.S.O. de mostrarán sus habilidades para escuchar a los demás.	Reflejar los sentimientos y clarificar la satisfacción.	Orientador	Durante el desarrollo de la unidad «escuchar» (en video) en la hora semana de tutoría.	· Reflejar los sentimientos correctamente a un nivel 3 o superior en la Escala de Carkhuff. · Reflejar la satisfacción y facilitar feed-back un mínimo de cuatro veces.	El orientador valida la competencia observando a los estudiantes durante las actividades y viendolos en la grabación del video.

FUENTE: Adaptado de Johnson y Whitfield (1991, p. 49).

502

502

BIBLIOGRAFÍA

Álvarez, V. (1993), «Ambitos temáticos y opciones metodológicas de la investigación en el campo de la orientación», *Revista de Orientación Educativa y Vocacional*, IV, 5, 11-23.

Álvarez, V. y Hernández, J. (1998), «El modelo de intervención por programas», *Revista de Investigación Educativa*, 16 (2), 79-123.

Álvarez, V.; Rodríguez, A.; Cruz, J.M. y Romero, S. (1995), «Investigación sobre orientación desde la universidad: entre la realidad y la utopía», *Revista de Investigación Educativa*, 26, 253-283.

American Counseling Association, (1995), *Code of ethics and standards of practice*, Alexandria, VA: Author.

Atkinson, D.; Furlong, M. y Janoff, D. (1978), «A four-component model for proactive accountability in school counseling», *School Counselor*, 26, 222-228.

Bartolomé, M. (1986), «Investigación-acción», en *Libro homenaje al Dr. Fernández Huerta*, Madrid, CSIC.

Borders, L.D. & Drury, S.M. (1992), «Comprenhensive school counseling programs: A review for policymakers and practitioners», *Journal of Counseling and Development*, 70 (4), 487-498.

Buendía, L.; Colas, P. y Hernández, F. (1998), *Métodos de investigación en psicopedagogía*, Madrid, McGraw-Hill/Interamericana.

Colas, P. (1990), «Aplicaciones y aportaciones de la investigación-acción participativa a los programas de intervención educativa», *Revista de Investigación Educativa*, 8, 16, 271-277.

Cook, T.D. y Reichardt, C.S. (1986), *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata.

Fairchild, T.N. (1986), «Time analysis: Accountability tool for counselors», *School Counselor*, 34, 36-43.

García Ramos, J.M. (1992), «Recursos metodológicos en la evaluación de programas», *Bordón*, 43, 4, 461-476.

Gelso, C.J. (1979), «Research in counseling: Methodological and professional issues», *The Counseling Psychologist*, 8 (3), 7-36.

Gysbers, N, & Henderson, P. (1994), *Developing and managing your school guidance program*, Alexandria, VA.: American Counseling Association.

Gysbers, N.; Hughey, K.; Starr, M. & Lapan, R. (1992), «Improving school guidance programs: A framework for program, personnel and results evaluation», *Journal of Counseling and Development*, 70 (5), 565-570.

Hansen, J.B. (1993), «Is educational reform through mandated accountability an oxymoron?», *Measurement and evaluation in counseling and development*, 26 (1), 11-21.

Heppner, P.; Kivlighan, D. & Wampold, B. (1993), *Research design in counseling*, Pacific Grove, CA: Brooks/Cole.

Johnson, S.K. & Whitfield, E.A. (1991), *Evaluating guidance programs*, Iowa city, IA: American College Testing.

BEST COPY AVAILABLE

Lambert, M.; Ogles, B. & Masters, K. (1992), «Choosing outcome assessment devices: An organizational and conceptual scheme», *Journal of Counseling and Development*, 70, 527-532.

Lázaro, A. (1983), «Tendencias temáticas en los estudios sobre orientación escolar en España», *Bordón*, 246, 17-27.

Lombana, J.H. (1985), «Guidance accountability: A new look at an old problem», *School Counselor*, 32, 340-346.

Maher, C.A. & Bennett, R.E. (1984), *Planning and evaluating education services*, Englewood Cliffs, NJ.: Prentice-Hall.

Marshall, C. & Rossman, G.B. (1989), *Designing qualitative research*, Newbury Park, CA: Sage.

Matuszek, P.A. (1981), «Program evaluation as consultation», en CONOLEY, J.C. (Ed.), *Consultation in schools: Theory, research, procedures*, New York: Academic Press.

Midkiff, R.M. Y Burke, J.P. (1987), «An action research strategy for selecting and conducting program evaluations», *Psychology in the Schools*, 24, 135-144.

Miguel, M. de (1985), «Evaluación de la investigación orientadora», en VVAA, *La orientación escolar ante la reforma de las enseñanzas medias y el empleo juvenil*, Madrid, 513-516.

Monette, D.; Sullivan, T. & Dejong, C. (1990), *Applied social research: Tool for the human services*, Fort Worth, TX: Holt, Rinehart & Wiston.

Patton, M.Q. (1990), *Qualitative evaluation and research methods*, Newbury Park, CA: Sage.

Pérez Juste, R. (1992), *Evaluación de programas de orientación*, Actas del V Seminario Iberoamericano de Orientación, Madrid, AEOEP.

Repetto, E. (1995), «La orientación como intervención psicopedagógica. Avances más relevantes en la investigación de algunos modelos», *Revista de Investigación Educativa*, 26, 233-251.

Rodríguez Espinar, S. (Coord.) (1998), «Orientación. Aspectos actuales y futuros» (nº Monográfico), *Revista de Investigación Educativa*, 16 (2).

Sanz Oro, R. (1993), «Evaluación de programas de orientación» (pp. 207-237), en Pérez Juste, R. et al. *Evaluación de programas y centros educativos*, Madrid, UNED.

Sanz Oro, R. (1998-3ª Ed. rev. y act.), *Evaluación de programas en orientación educativa*, Madrid, Pirámide.

Sanz Oro, R. (1999), *Los departamentos de orientación en educación secundaria: roles y funciones*, Barcelona, Cedecs.

Sexton, T.; Whiston, S.; Bleuer, J. & Walz, G. (1997), *Integrating outcome research into counseling practice and training*, Alexandria, VA: American Counseling Association.

Tejedor, F.J. (1994), «Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo», *Revista de Investigación Educativa*, 23, 93-127.

Walz, G. & Bleuer, J. (1993), *Counselor efficacy. Assessing and using counseling outcomes research*, ERIC/Counseling and Personnel Services Clearinghouse, School of Education, The University of Michigan, Ann Arbor.

Weis, C.H. (1991), *Investigación evaluativa*, Méjico, Trillas.

Whiston, S.C. & Sexton, Th.L. (1998), «A review of school counseling outcome research: Implications for practice», *Journal of Counseling and Development*, 76 (4), 412-426.

Whiston, S.C. (1996), «Accountability through action research: Research methods for practitioners», *Journal of Counseling and Development*, 74, 616-623.

Worzbyt, J.C. & O'rourke, K. (1989), *Elementary school counseling: A blueprint for today and tomorrow*, Muncie, Indiana: Accelerated Development Inc.

505

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 521-537

LA EVALUACIÓN DE LA FORMACIÓN OCUPACIONAL Y CONTINUA

J.M. Jornet, J.M. Suárez y Mª J. Perales[1]

1. INTRODUCCIÓN

La Evaluación de los Programas de Formación para el Trabajo ha concitado un gran interés en los últimos años, tanto desde una perspectiva sociopolítica, como desde la gestión y organización de los mismos. La estructuración y evaluación de la Formación Profesional (Cabrera, 1987), en sus diversas modalidades, desde la Inicial —reglada— a la no reglada, ha constituido un centro de interés que progresivamente se ha ido instaurando también en nuestro país. El motivo central es que se ha constituido como uno de los ejes centrales de dinamización del empleo y de mejora de la competitividad empresarial a nivel europeo. Los fondos aportados para la Formación, las expectativas creadas en cuanto a su necesidad y consecuencias sociales, han definido en gran medida las soluciones evaluativas que se han ido desarrollando. Por este motivo, hemos estimado conveniente, antes de entrar en el ámbito propiamente dicho de la Evaluación, revisar algunas características de este tipo de Programas, dado que su Evaluación se orienta y/o condiciona por ellos.

La estructura del presente artículo es la siguiente: a) Algunas notas características acerca de la Formación Ocupacional y Continua, b) Una aproximación a una tipología de Planes de Evaluación, y c) Reflexiones en torno a Estándares para la Acreditación y/o Certificación.

2. UNAS NOTAS ACERCA DE LA FORMACIÓN OCUPACIONAL Y CONTINUA

Para realizar la revisión señalada anteriormente, tomamos como referencia tres aspectos de la Formación: a) Sus *Finalidades*, b) *Modalidades*, y c) *Condiciones y contexto de desarrollo.*

---

1 Dpto. MIDE de la Universitat de València. Fac. Filosofía y CC. De la Educación. Avda. Blasco Ibáñez, 30. Tél. 963864430. e-mail: jornet@uv.es, rodrigus@uv.es, perales@uv.es

506

Como indicamos en la introducción, la Formación para el trabajo cobra relevancia en los últimos años, como un elemento de política activa de dinamización del Empleo. Así, en el marco de las políticas socioeconómicas tanto europeas como nacionales, el rol que se le asigna a la Formación es el de mejora de los Recursos Humanos y el incremento de la competitividad empresarial. De este modo, la Formación se asume como un medio para aumentar las posibilidades de empleo de las personas y facilitar su adaptabilidad a los cambios en los procesos productivos.

Dentro del ámbito de este tipo de *Formación*, la *Ocupacional* está orientada a mejorar la inserción sociolaboral de las personas; y, si bien está dirigida a los desempleados[2] en general, una parte de sus Programas se especifican hacia colectivos con riesgo de exclusión del mercado laboral (parados de larga duración, jóvenes sin formación profesional y que no han disfrutado de un primer empleo o mujeres separadas del mecado laboral durante un largo periodo de tiempo, por ejemplo). No obstante, los tipos de Programas son muy diversos y pueden darse desde Programas de Iniciación en una Ocupación, integrando la formación en aprendizajes básicos, hasta otros de una elevada especialización.

Por su parte, la *Formación Continua* se dirige a personas con empleo. Su función es mejorar pues la capacitación personal y, con ello, favorecer la promoción de los trabajadores o adaptar los Recursos Humanos a los cambios de los procesos de producción que se dan —o se prevén— en un sector o en una empresa determinada. Por ello, el rango de Programas, aunque también muy amplio, se suele circunscribir a un nivel de especialización mayor.

En cuanto a las *Modalidades* en que se produce la Formación, ciertamente no difieren de las tres básicas: *Presencial, Abierta y a Distancia*. Los Programas más frecuentes son los Presenciales y, a continuación, la formación que integra módulos de formación convencionales y atención tutorial en el puesto de trabajo y, finalmente, los Programas a Distancia. No obstante, esta distribución se produce fundamentalmente por el nivel de desarrollo del sector de la Formación, más que por su adecuación a las características de los colectivos a que se dirige, sus necesidades, el objeto de la Formación, etc...

Finalmente, las *Condiciones y contexto de desarrollo* de la Formación se pueden sintetizar en los siguientes puntos (Villanueva y Catalá, 1997; Jornet et. al., 1998):

- El estímulo fundamental para el desarrollo de este tipo de Programas se produce desde la orientación política, tanto desde la administración europea como desde los usos concomitantes que realizan las administraciones nacionales y regionales. La financiación, por tanto, también es fundamentalmente pública, por lo que es muy importante la necesidad de controlar la adecuación de los usos de los Fondos Públicos, los efectos de este tipo de Programas, etc... Además, esta misma preocupación, también se produce cuando es la propia empresa la que impulsa y financia la formación.

---

2 Para favorecer el acceso a un primer empleo o su reingreso en el mercado laboral, bien a través de una mayor especialización en su formación, bien iniciando a la persona en otra Ocupación, facilitando con ello el cambio profesional.

507

- El hecho de que la financiación provenga de fondos públicos, implica que la oferta de Formación se estructure en función de una Programación periódica, generalmente de carácter anual, y que se vea condicionada por ella en sus líneas de actuación, tiempos de realización, etc...
- Por otra parte, se produce una flexibilidad curricular extrema, de forma que ante una misma denominación de Programa pueden desarrollarse curricula de diversa longitud, contenido, finalidad, etc... Así, aunque existen referentes curriculares desarrollados por las Administraciones, éstos no son vinculantes; pudiendo recibir financiación por parte de la misma Administración soluciones curriculares muy diferentes para una misma denominación.
- La heterogeneidad aludida respecto a los Programas, también se observa con relación a las entidades e instituciones de Formación. Este sector es atendido, tanto por la administración pública, como por la actuación de agentes sociales (asociaciones empresariales y sindicales), o por entidades privadas (academias, consultorías, etc...).
- Por otra parte, las entidades que realizan Formación son normalmente organizaciones dedicadas a la gestión y, en su estructura, en raras ocasiones se identifican especialistas en Pedagogía. Asimismo, el corpus docente tampoco es estable y generalmente, salvo en casos de Consultoras que actúan en ámbitos muy específicos de la Formación Continua (Mejora Organizacional, Marketing, Dirección, etc...), está compuesto por profesionales que ocasionalmente actúan como docentes. Lo más frecuente es que la Formación Ocupacional sea impartida por profesionales que, a su vez, también están temporalmente en paro en su profesión habitual.
- Los colectivos a que se dirigen ambos tipos de Formación son también muy diversos. No obstante, esta realidad no se suele traducir en una metodología didáctica especializada, diferencial, para cada uno de ellos.

Este conjunto de características condiciona ineludiblemente los tipos de Evaluación que se han ido desarrollando. Fundamentalmente, son los intereses de los promotores de la Formación y, en segundo lugar, de los Gestores los que han ido orientando el *objeto a evaluar* (qué se evalúa en la Formación) y su *finalidad* (para qué), así como la identificación de *indicadores*.

3. TIPOS DE PLANES DE EVALUACIÓN

La gran evolución que, según se ha indicado, ha experimentado la evaluación de la Formación Ocupacional y Continua no ha ido acompañada de un desarrollo sistemático de la propia evaluación. El contexto de la Formación Ocupacional y Continua es muy peculiar y tiene varias características que explican en parte esta situación:

- La formación y su evaluación siguen un sistema de gestión similar al de otros elementos de la empresa, de modo que la persona que se encarga de ellas frecuentemente carece de formación pedagógica y en evaluación.

508

- Por ello, las prácticas evaluativas no suelen estar sistematizadas ni responder a un modelo de evaluación previo.
- El carácter interno de muchas de estas evaluaciones hace muy limitada su difusión, siendo escasas las ocasiones en que se publican los resultados y/o los planes utilizados.

No obstante, a partir del análisis de una serie de dimensiones básicas y de la revisión de las propuestas y experiencias de Evaluación de la Formación Ocupacional y Continua publicadas en nuestro contexto, se puede plantear una *propuesta de Tipología de Planes de Evaluación* de este tipo de formación (Jornet, 2000; Perales, 2000), que sirva para estructurar estas prácticas evaluativas, y que queda totalmente abierta a sugerencias y revisiones, conforme se vayan difundiendo nuevas propuestas y experiencias en este ámbito.

Esta tipología parte del análisis del concepto de *calidad* que tiene como referente cada plan de evaluación. Es cierto que detrás de cada plan de evaluación hay un concepto de calidad que lo orienta, pero éste no siempre está explícito y, cuando lo está, no siempre se corresponde con las prácticas evaluativas efectivamente implementadas, de modo que se suele detectar una gran distancia entre la *utilidad declarada* de un plan de evaluación, y su *utilidad real*, lo cual afecta negativamente a la propia *credibilidad* de los procesos evaluativos (Jornet, 2000).

Por ello, el análisis se hace en función de la operativización del concepto de calidad, de cómo se concreta en la práctica evaluativa, analizado desde tres dimensiones:

- *Globalidad del análisis*. Se refiere a si el concepto de calidad se operativiza mediante enfoques holistas y globalizadores (macro analíticos) o mediante planteamientos analíticos y detallados (micro analítico).
- *Focalización de la evaluación*. Esta dimensión está definida por el objeto real de la evaluación. En sus extremos, puede ser:
  - — *Extrínseco a la formación*, cuando se valora la calidad de la formación a partir de sus consecuencias finales e indirectas (como la inserción, la promoción, la satisfacción...).
  - — *O intrínseco*, cuando se valora la calidad de la formación en sí misma (adecuación de la organización, programación, actuación docente, calidad de los materiales e infraestructuras...).
- *Tipo de información* sobre la que se sustenta la idea de calidad. Esta tercera dimensión hace referencia al tipo de indicadores sobre los que se basa la evaluación de la calidad, definiéndose en sus extremos por:
  - — *Informaciones directas*, que nacen directamente del análisis específico del objeto.
  - — *Informaciones indirectas*, que se refieren al objeto, pero no lo analizan directamente.

Además de estas dimensiones, derivadas del concepto de calidad que subyace en el plan de evaluación, la tipología que se presenta analiza tres más, como componentes fundamentales del proceso de evaluación (Jornet, Suárez y Belloch, 1998):

- Objeto de evaluación.
- Finalidad.
- Control.

El esquema de dimensiones sobre el que se construyen los distintos planes de evaluación de la Formación Ocupacional y Continua puede sintetizarse en la siguiente tabla:

TABLA 1
DIMENSIONES DE CLASIFICACIÓN DE LOS PLANES DE EVALUACIÓN

Calidad			Otros componentes		
Globalidad del análisis	**Focalización de la evaluación**	**Tipo de información**	**Objeto de la evaluación**	**Finalidad**	**Control**
Macro analítico	Extrínseca	Inf. Directas	(ver Tabla 2)	Rendición de cuentas	Externo
Micro analítico	Intrínseca	Inf. Indirectas		Mejora	Interno

De acuerdo con estas dimensiones, los enfoques de evaluación de la Formación Ocupacional y Continua se pueden agrupar en cuatro grandes tipos:

- Planes de Evaluación de Programas de Formación Ocupacional y Continua como herramienta de acción sociopolítica.
- Planes de Evaluación Institucional.
- Planes de Evaluación de Programas.
  - — Enfoque instructivo / educativo.
  - — Enfoque global.
- Planes de Evaluación de Recursos Humanos y Materiales.

Planes de Evaluación de Programas de Formación Ocupacional y Continua como herramienta de acción sociopolítica

El *objeto* de evaluación de estos planes son las consecuencias sociales del Plan de Formación Ocupacional o Continua. Valoran la consecución de los objetivos socioeconómicos descritos para el plan de formación, entendiendo la formación como un medio, pero no un fin en sí misma. Los indicadores que se utilizan hacen referencia al impacto socioeconómico de la formación, lo cual se traduce en cuestiones como la inserción, la promoción, la mejora del rendimiento en el puesto de trabajo, el incremento de la productividad, la mejora del clima de trabajo, etc... Aunque menos frecuente, también existe dentro de esta tipología un conjunto de trabajos que se orienta a

la valoración sociopolítica de las acciones y macroprogramas financiados por las instituciones, particularmente los Estados; éstos se centran fundamentalmente sobre componentes económicos y sociales, como por ejemplo, las inversiones realizadas, la implicación y participación de los agentes sociales en las decisiones e implementación de los programas, las facilidades de los diferentes implicados en cuanto a su desarrollo, etc. (Keep y Mayhen, 1999).

Las principales características de este tipo de Planes de Evaluación son:

- El concepto de *calidad* que sustenta estos planes suele traducirse en estudios:
  — macro analíticos (trabajan sobre grandes muestras, analizando dimensiones globales).
  — centrados en elementos extrínsecos a la propia formación.
- La *finalidad* de estos estudios, dependiendo de dónde surja la iniciativa y de cómo sean diseñados puede recorrer desde la rendición de cuentas (enfoque sumativo) hasta la orientación para la mejora o la innovación (enfoque formativo), pasando por todos los planteamientos mixtos. Sin embargo, lo habitual es que se trate de estudios de rendición de cuentas, con un control externo.

La mayoría de los estudios de evaluación realizados sobre la Formación Ocupacional, tanto en el contexto europeo en general, como en concreto en el caso español, tienen esta orientación. En el caso de la Formación Continua, la dificultad de operativización del indicador de impacto en medidas observables (mejora del rendimiento en el trabajo, aumento de la productividad, mejora del clima de trabajo, mejora de la calidad del trabajo...) ha provocado que frecuentemente los estudios evaluativos se sustenten en medidas indirectas de satisfacción de las personas, o de valoración global de los efectos (a través de las opiniones de los supervisores / responsables).

TABLA 2
CARACTERÍSTICAS CLAVE DE LOS PLANES DE EVALUACIÓN SOCIOPOLÍTICA

Planes de Evaluación Sociopolítica	
• Objeto de evaluación	Consecuencias sociales del programa
• Operativización concepto de calidad	Focalización extrínseca. Estudios macro analíticos.
• Finalidad	Habitualmente, rendición de cuentas
• Control	Habitualmente, externo
Es el tipo de evaluación más frecuente en Formación Ocupacional.	

Los estudios publicados sobre evaluación de este tipo de programas formativos por la OCDE (1991) y el CEDEFOP (1997) parten de la base de planes de evaluación construidos sobre indicadores como:

- La idoneidad
  Se refiere a la relación entre los objetivos de la formación y las necesidades del mercado de trabajo.
  - *«¿está el programa correctamente planteado?*
  - *¿cumple sus objetivos externos (a la formación)?*
  - *¿está motivado por una demanda?*
  - *¿están satisfechos los alumnos?»* (CEDEFOP, 1997: 40)

- La eficacia
  Estudia la relación entre los objetivos y los resultados.
  - *«¿cumple el programa correctamente con lo que se supone que ha de cumplir, esto es, produce mano de obra cualificada con la calidad y la cantidad que exigen los mercados de trabajo?*
  - *¿satisface sus objetivos internos, sobre todo en términos de cantidad y calidad de los resultados?*
  - *¿los participantes terminan por encontrar un empleo y mejorar su productividad y sus ingresos?*
  - *¿alcanza el programa a los destinatarios que pretende (jóvenes, parados de larga duración, personas sin cualificación inicial...)?»* (CEDEFOP, 1997: 40).

- La rentabilidad (o eficiencia)
  Estudia la relación entre la inversión y los resultados:
  - *«¿hace el programa el mejor uso posible de los recursos puestos a su disposición?*
  - *¿sería posible reducir costes y/o conseguir recursos adicionales?»* (CEDEFOP, 1997: 40).

Planes de evaluación institucional

Cuando la evaluación de la Formación Ocupacional y Continua se realiza mediante planes de evaluación institucional toma como *objeto* de análisis una *entidad*. El énfasis, en este caso, se sitúa en el análisis de la organización, sus componentes de gestión, sus infraestructuras, etc.

La *finalidad* del proceso de evaluación determina en gran medida los distintos componentes de la misma, pudiéndose identificar dos polos opuestos:

- *«Procesos como los de homologación, acreditación o certificación de Centros, se sitúan cercanos a usos sumativos o mixtos de la evaluación, y suelen realizarlos evaluadores externos»* (Perales, 2000: 140):
  — Se utilizan modelos como el EFQM, o basados en la ISO 9000, o bien protocolos propios diseñados según los objetivos del proceso de homologación.

— Pueden responder a la demanda de la entidad financiadora (administración generalmente), o a una política de mercado, basada en ofrecer una imagen de calidad a partir de la certificación.

- *«En el otro extremo, los procesos orientados por las propias entidades como actuaciones de autorrevisión están orientados hacia la mejora e innovación institucional, y son llevados adelante por equipos mixtos o por evaluadores internos»* (Perales, 2000: 140):
  — El Modelo EFQM puede adecuarse perfectamente a las expectativas y necesidades de estas entidades, aunque habitualmente utilizan diseños de evaluación emergentes, construidos desde la propia entidad.

Entre estos casos extremos se sitúa toda la riqueza y la variedad de los distintos procesos evaluativos, con los matices derivados de su finalidad específica, y de cómo haya sido operativizada.

TABLA 3
CARACTERÍSTICAS CLAVE DE LOS PLANES DE EVALUACIÓN INSTITUCIONAL

Planes de Evaluación Institucional		
• Objeto de evaluación	La entidad	
• Operativización concepto de calidad	Focalización intrínseca. Nivel medio de globalidad.	
• Finalidad	Polo 1	Polo 2
	Homologación Acred / Certificación	Autorrevisión para la mejora
• Control	Habitualmente, externo	Interno

Planes de evaluación de programas

El tercer tipo de Planes de Evaluación se define en función de los Programas, e intenta valorar si se han conseguido los objetivos pretendidos. Constituyen en sí mismas aproximaciones complejas en las que debe adaptarse específicamente el Plan de Evaluación a las características y finalidades de los programas (Carballo, 1991). En función de la amplitud de este concepto de *eficacia* se podrían diferenciar dos subtipos de enfoques (Perales, 2000):

- *Enfoque instructivo / educativo.*

  Cuando la eficacia se entiende como la consecución de los objetivos formativos (niveles de conocimiento, ejecución, competencias, habilidades personales...) la evaluación del programa se circunscribe a valorar este tipo de objetivos.

Se trata de una evaluación de programas similar a la que se puede dar en cualquier otro nivel educativo. Se trata de evaluaciones que:

- Requieren una descripción bastante precisa de los distintos *componentes* que tienen que ver con el programa evaluado (diseño, metodología, tipos de actividades, materiales necesarios, infraestructuras...).
- Por tanto, tienen un *concepto de calidad* operativizado en diseños micro analítico, con focalización intrínseca.
- Se diseñan, por lo demás, según *la finalidad y el objeto* concreto a que se apliquen. Así, pueden plantearse con finalidad formativa, sumativa o mixta, con control interno o externo, y utilizando medidas directas o indirectas.
- La amplia experiencia en este tipo de planes de evaluación en el contexto de la educación reglada contrasta con su *escasa implantación* en el ámbito de la Formación Ocupacional y Continua, pues en este caso las evaluaciones no se suelen realizar desde una perspectiva pedagógica. Sin embargo, es esta perspectiva pedagógica la que puede permitir identificar claves para la mejora de la calidad los programas de formación, desde un punto de vista educativo, influyendo directamente en los objetivos planteados, en los materiales, en la metodología didáctica, en los procesos de evaluación de los alumnos... Es más, este tipo de evaluación puede facilitar el proceso de certificación / acreditación, al ofrecer referentes en los que basarlo y establecer equivalencias entre programas.
- No obstante, como *ejemplos* se pueden citar las experiencias del Programa francés de evaluación de los *Atéliers de Pédagogie Personalisé* de Bretaña (DRTEFP, 1994), y la evaluación del Programa TETRA (Jornet y Suárez —Coords.—, 1999), referido a Formación Continua mediante teleformación.

- *Enfoque global*

La consideración de la Formación Ocupacional y Continua como políticas activas de empleo se traduce en un concepto más amplio de eficacia, donde a los objetivos formativos se añade la valoración de la consecución de otros objetivos sociolaborales implicados (inserción laboral, promoción en el empleo, mejora del rendimiento en el trabajo...). En este caso, los planes de evaluación son más amplios y abarcan ambos niveles. Se trata, por tanto, de evaluaciones que:

- Requieren una definición clara del *rol que juegan las distintas variables* en el plan de evaluación, siendo para ello muy eficaces modelos como el CIPP, de Stufflebeam.
- La operativización del *concepto de calidad* en estos planes está determinada por ese enfoque global, mixto, de modo que combinan la focalización extrínseca e intrínseca (al abordar tanto elementos formativos como referentes a sus consecuencias), y se sitúan en niveles de globalidad intermedios (ni tan

micro analíticos como la evaluación puramente educativa, ni las macro analíticos como los estudios sociopolíticos).

- Como en el caso de la evaluación más educativa, el diseño concreto de los estudios dependerá de *la finalidad y el objeto* concreto a que se apliquen.
- El *acceso* a este tipo de evaluaciones es difícil, pues las que se realizan, además de ser poco frecuentes, suelen ser de carácter interno, a partir de la iniciativa del equipo pedagógico de las entidades (aquéllas que cuentan con él), y no se suelen difundir. No obstante, se pueden citar los siguientes *ejemplos*:
  - — Las propuestas de Tejada (1992) y de Martínez Piñeiro (1995), referidas a Formación Ocupacional.
  - — La evaluación del Programa francés «FNE-Cadres» (DRTEFP, 1997).
  - — La Evaluación de la Formación Ocupacional y Continua realizada en la Comunidad Valenciana (Jornet y Suárez, 1997; Perales, 2000), que en una de sus dimensiones puede considerarse un plan de evaluación global de este tipo.

TABLA 4
CARACTERÍSTICAS CLAVE DE LOS PLANES DE EVALUACIÓN DE PROGRAMAS

Planes de Evaluación de Programas		
• Objeto de evaluación	Programa formativo	Programa global
• Operativización concepto de calidad • Focalización • Globalidad	 Intrínseca Nivel micro analítico	 Mixta Nivel intermedio
• Finalidad	Rendición de cuentas y/o mejora	
• Control	Externo y/o interno	
Suelen ser experiencias poco difundidas, tanto sus modelos como sus resultados.		

Planes de evaluación de Recursos Humanos y/o Materiales

En este último tipo de planes de evaluación se recogen las propuestas que están dirigidas a elementos concretos implicados en las Acciones Formativas (como puede ser la evaluación del profesorado, de materiales didácticos, de las prácticas...) pero que carecen de la visión sistémica que define los tipos anteriores.

No se trata, por tanto, de una categoría real de clasificación, sino de un recurso para recoger los numerosos estudios evaluativos de este tipo que se producen en el ámbito de la Formación Ocupacional y Continua, respondiendo a un interés específico que no se encuadra en un estudio evaluativo global.

La variedad de diseño de estos planes es casi infinita, pues dependerá de los objetos específicos de evaluación y de las finalidades del proceso. Los más habituales, no obstante, son quizá los referidos a evaluación de materiales (sobre todo en los programas que se imparten a distancia) y los de evaluación del profesorado, a partir de las opiniones de los alumnos (realizados en muchas entidades de formación, como proceso interno de control y mejora).

Aunque las entidades suelen realizar evaluación de profesorado, como se ha indicado, se trata de estudios internos, que no son difundidos o que tienen un ámbito de difusión limitado o muy especializado. Por tanto, como ejemplo se puede citar la valoración de materiales que se desarrollan al amparo de Proyectos Europeos (vgr. Leonardo da Vinci, etc...).

4. EL DESARROLLO DE ESTÁNDARES PARA LA EVALUACIÓN, ACREDITACIÓN Y CERTIFICACIÓN

La flexibilidad y versatilidad de la Formación para el Empleo ha puesto de manifiesto la necesidad de establecer referentes que permitan analizar los contenidos de los Programas, las competencias adquiridas por las personas, o la capacidad de las instituciones para impartir este tipo de actividades de formación. Por tanto, se están desarrollando diversas líneas de trabajo, en las que se profundiza en uno de los usos de la evaluación actualmente más demandados: la Acreditación y/o Certificación.

En este contexto, se orienta la evaluación hacia el marco del desarrollo de ***Estándares***, que pueden ser definidos desde dos acepciones distintas, si bien concitan un gran interés actualmente para su aplicación en cualquier nivel educativo (De la Orden, 2000; Mateo, 2000, Pérez Juste, 2000; Jornet y Suárez, 2000):

a) El estándar como protocolo de actuación para el desarrollo de evaluaciones, y que se define generalmente por normas de procedimiento, y
b) El estándar como síntesis de criterios de calidad.

No obstante, estas acepciones no son unívocas, y pueden presentar asimismo diferentes matices, dependiendo de la aplicación concreta en la que se alude al *Estándar*.

Se pueden identificar tres grandes categorías de estándares, en función del objeto evaluación:

- Estándares de Certificación / Acreditación de Instituciones.
- Estándares de Certificación / Acreditación de Programas.
- Estándares de Certificación / Acreditación personal.

Estándares de Certificación / Acreditación de Instituciones

Como se indicaba anteriormente, la preocupación por la calidad iniciada en el ámbito empresarial se ha traducido en un interés por la calidad de la propia institución formativa, y por demostrarla en virtud de una certificación que así lo acredite.

Este proceso de certificación, que en Europa tiene ya una larga tradición, está llegando a nuestro país en los últimos años. Los estándares que se utilizan se basan en dos grandes modelos: las Normas ISO y el MEGC. Ambos, si bien se han desarrollado en el ámbito productivo y/o empresarial, se han aplicado a Instituciones Educativas. No obstante, el Modelo Europeo de Gestión de Calidad, basado en el EFQM, está teniendo una mayor aceptación actual, porque se adapta mejor al análisis de las organizaciones educativas, al estar diseñado directamente para analizar organizaciones.

Los estándares, en ambos casos, se centran en elementos de la organización, fundamentalmente en los que podríamos señalar como indicadores de la *Calidad del Servicio* que aporta la institución. Estos procesos evaluativos, independientemente del modelo que utilicen, promueven la calidad en las instituciones, al desarrollar una cultura de evaluación, a través de la autorrevisión. Sin embargo, aunque estas revisiones se realizan atendiendo al ámbito formativo que va a desarrollar cada institución, lo cierto es que no se incide específicamente en la evaluación de los programas, de modo que en algunos casos la reflexión sobre los componentes curriculares y los resultados de los mismos puede ser mejorable.

Por otro lado, el concepto de estándar que subyace a ambos casos se basa en un protocolo de actuación, de modo que no existen criterios explícitos de calidad. La certificación, por tanto, supone que la entidad se ha implicado en un proceso de autorrevisión, tendente a la mejora de su calidad, pero no supone un nivel mínimo estandarizado.

Asimismo, en nuestro país, se pueden identificar experiencias asumibles dentro del marco de la Acreditación de Centros. Por ejemplo, en la Comunidad Valenciana, la Consellería de Trabajo y Asuntos Sociales puso en marcha en el año 1997 un proceso de Homologación de Centros, como requisito para la financiación de Programas de Formación Ocupacional. El sistema se basa en el análisis que realiza la CTAS de diversos componentes de los Centros. Esta revisión, se orienta sobre un protocolo en el que se establecen los requisitos que deben cumplirse para impartir Programas de determinadas Familias Profesionales. Así, la homologación se realiza para un sector específico. El sistema se completa con la supervisión que realizan los Técnicos de Enlace (supervisores de la administración) durante el desarrollo de los Programas financiados. En este caso, hay que señalar que aunque la homologación no se presenta como un sistema de Estándares para la Acreditación, ciertamente lo es y los componentes que se revisan son muy específicos: infraestructuras, medios didácticos, etc... Sin embargo, como en los casos señalados anteriormente, los requerimientos vinculados a los diseños curriculares son escasos y son frecuentemente olvidados o desatendidos.

Actualmente esta línea se completa con la aplicación del Modelo Europeo de Gestión de Calidad. De este modo, los centros propios del sistema público están participando en un sistema de autorrevisión, orientado desde el gobierno autonómico, para mejorar la calidad del servicio que imparten.

Estándares para la Certificación /Acreditación de Programas

Se refieren a los criterios que permiten establecer el grado de calidad de un programa de formación, identificando los niveles de competencia que permite adquirir a las

517

personas que participan en él. Por tanto, que un programa esté acreditado debería suponer:

- Que tiene un diseño curricular correcto (bien adaptado a la ocupación, con una programación optimizada, con una metodología didáctica adecuada...)
- Que se desarrolla de un modo adecuado
- Que permite adquirir las competencias para las que está diseñado
- Que las personas que lo siguen y superan con aprovechamiento demuestran haberlas adquirido.

Este planteamiento requeriría el uso de una evaluación pormenorizada, en la que se controlaran variables de diferente rol (de contexto, de entrada, de proceso y de producto), y que permitiese la optimización del programa (Pérez Carbonell, 1999).

Sin embargo, en el sector de la Formación Ocupacional y Continua son muy escasos los programas que han estado sujetos a este tipo de revisión. Más específicamente, se suele analizar el proceso de implementación de un programa, y, en ocasiones, las características de su diseño curricular, pero muy raramente se evalúa la competencia adquirida por las personas. Desde este punto de vista, la certificación o acreditación de programas se suele circunscribir a procesos de evaluación de Calidad del Servicio, sin abordar la efectividad real del programa.

La heterogeneidad del ámbito de la Formación para el Empleo condiciona en buena medida que la evaluación de los programas sea de este tipo. La evaluación es una actividad innovadora, y está iniciándose ahora, a través de los elementos que, de momento, son más accesibles. El objetivo, no obstante, debe ser avanzar hacia procesos evaluativos que aborden los programas en todas sus dimensiones, incluida la valoración de las competencias que permiten adquirir.

Como en el caso anterior, en el ámbito de la evaluación de programas se están desarrollando propuestas que tienden a incidir en la idea de estándares como protocolo de actuación evaluativa, más que como criterios específicos de calidad y eficacia a demostrar por un programa. En este sentido, los estándares de orientación al rendimiento para la Evaluación de Programas (OCDE-PUMA/PAC, 1999) tienen por objeto:

- crear las condiciones previas favorables para el desarrollo de la evaluación,
- desarrollar un marco efectivo de trabajo,
- incidir sobre aspectos prácticos del proceso de formación y evaluación, y
- crear una cultura de evaluación orientada a la mejora de la calidad del programa.

Estándares de Certificación / Acreditación personal

La movilidad laboral se ha convertido en una constante en la evolución profesional de las personas, y está exigiendo sistemas que permitan certificar las competencias adquiridas, sea a través de programas de formación y/o a partir de su experiencia profesional.

La *certificación de la formación* recibida suele basarse en documentos emitidos por las entidades formadoras, sean regladas o no regladas. En el ámbito de la Formación Ocupacional y Continua en ocasiones su credibilidad está cuestionada, pues los certificados no acreditan la competencia, sino únicamente la participación en los cursos. Esto es así dado que generalmente las condiciones en que se desarrollan estos cursos hacen que se den las máximas facilidades a los alumnos para que participen en ellos[3]. Así, aunque se incluyen evaluaciones de los alumnos, generalmente no se utilizan como elemento para diferenciar entre aquellos alumnos que han seguido con aprovechamiento el curso, y aquellos que no lo han hecho.

Un problema más arduo lo constituye la *certificación de la experiencia profesional*. El objetivo sería poder certificar un determinado grado de competencia en función del trabajo que haya realizado cada persona. Sin embargo, los curricula no son elementos suficientemente explícitos y fiables para este cometido, pues frecuentemente bajo una misma denominación de puesto de trabajo, las funciones que ha desarrollado la persona son muy diferentes de una empresa a otra; además, también es frecuente la práctica en la que el trabajador se integra oficialmente en una categoría —por conveniencia de la empresa—, mientras que realiza otras funciones.

Estas disfunciones son excesivamente frecuentes y perjudican tanto a los trabajadores como a las empresas en los procesos de selección de personal. En esta línea se están desarrollando diversas propuestas orientadas a la certificación de los componentes de formación: la experiencia a lo largo de la vida, los Programas, etc... Un ejemplo que ha concitado gran interés es el caso de los NVQs (National Vocational Qualification) británicos, que tienen como objetivo (Training Agency, 1988):

- Incrementar la responsabilidad de los empleados,
- Aportar y promover formación para los jóvenes, y
- Ayudar a los desempleados.

Los Estándares que se proponen desde los NVQs pretenden establecer criterios para describir de forma fiable la competencia de los trabajadores. A la base se sitúa la posibilidad de poder establecer comparaciones respecto a la competencia, basados en la integración de diversos componentes formativos, que se establecen como equivalentes. Por lo tanto, estos estándares deben ser:

- relevantes para el trabajo,
- validados a nivel nacional, y
- establecidos por agentes sociales.

Sin embargo, aunque son múltiples los esfuerzos que se están desarrollando a nivel europeo (a través de los gobiernos, las convocatorias de investigación, desarrollo y for-

3 En la mayor parte de Programas financiados con fondos públicos (Europeos, Estatales y Autonómicos) se identifica como requisito el número mínimo de alumnos que debe seguir el Programa. En muchas ocasiones las entidades formativas dependen en exceso de que los alumnos se mantengan en el Programa y lo concluyan, pues de ello depende que se dé finalmente la financiación completa o se minore.

mación, etc...), el camino aquí se vislumbra extenso y plagado de dificultades. Las orientaciones que se están apoyando pasan necesariamente por un planteamiento metodológico que facilite la descripción de las características de la formación y de las competencias adquiridas.

A modo de Conclusión

La dinámica de la evaluación, por tanto, se está desarrollando ampliamente en el ámbito de la Formación Ocupacional y Continua. No obstante, es conveniente revisar los procesos de evaluación y promover su mejora en cuanto a su credibilidad y utilidad.

Algunos de los puntos sobre los que se podría sustentar esta pretensión son:

- *Integrar la mejora de los procesos evaluativos en una renovación global de la Formación Ocupacional y Continua.* Como en cualquier otro ámbito, hay que tener presente que la evaluación se puede constituir en un instrumento para la innovación. Sin embargo, como se ha demostrado en el sistema educativo, hay que evitar generar excesivas expectativas respecto a la evaluación, pues una verdadera mejora de los procesos formativos requiere de un verdadero compromiso, tanto desde la propia institución implicada como desde el contexto sociopolítico en que se inserta.
- *«Pedagogizar» la Formación y la Evaluación.* Tanto a nivel europeo como a nivel nacional, se está asumiendo ya como un criterio de calidad de la formación la atención explícita a diversos componentes pedagógicos de los mismos, más allá de meras cuestiones organizativas y de gestión.
- *Asumir la necesidad de determinar Estándares de Ejecución en los Programas.* Estos estándares tendrían un efecto multiplicador en cuanto a su utilidad, dado que, además de su aplicación para la evaluación de programas, permitirían establecer criterios claros de competencia para las certificaciones de las personas e instituciones.
- *Desarrollar Protocolos de Actuación bien determinados para el desarrollo de las Evaluaciones (Normas).* Estos protocolos deben atender a todo el proceso de evaluación, desde su contratación/encargo —en los que deben establecerse todos los componentes implicados en la evaluación, desde sus formas de realización al uso de la información—, pasando por su planificación y desarrollo, hasta la difusión y uso de la información resultante.
- *Organizar la actuación evaluadora sobre modelos de profesionalización.* Todo el proceso debe estar presidido por la idea de profesionalización de la evaluación, y sujeto a normas éticas que debemos ir consensuando entre todos (González, Pérez y Perales, 2000).

El ámbito de la Formación Ocupacional y Continua necesita y requiere no sólo de las aportaciones de la evaluación, sino también de toda una línea de trabajo vinculada a la pedagogía que va a contribuir efectivamente a la mejora de su calidad.

520

BIBLIOGRAFÍA

Cabrera F.A. (1987). *La investigación evaluativa en Educación. Técnicas de Evaluación y Seguimiento de Programas de Formación Profesional.* Barcelona: Fundación Largo Caballero.

Carballo, R. (1991). Introducción a la Evaluación de Programas de Acción Social: Decisiones para su realización. *Revista Complutense de Educación, 2 (1),* págs. 111-126.

Cedefop (comp) (1997). *Evaluación de programas europeos de formación, empleo y recursos humanos.* Conferencia de Atenas, 1995. Salónica: CEDEFOP.

De la Orden, A. (2000). Criterios para la elaboración de estándares en evaluación educativa. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas.* Valencia, Marzo (en prensa).

Drtefp Bretagne (1994). *Modèle d'Évaluations des APP de Bretagne.* Documento no publicado.

DRTEFP Ile de France (1997). *FNE-Cadres. Appel a propositions 1998.* Documento no publicado.

González J.; Pérez Carbonell, A. y Perales M.J. (2000). Componentes de formación y profesionalización en la evaluación. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas.* Valencia, Marzo (en prensa).

Jornet, J. (2000). *Evaluación del Profesorado Universitario.* Ponencia presentada en el *Seminario de Evaluación Educativa* de la ASOCIACIÓN NACIONAL DE INSPECTORES DE ENSEÑANZA SECUNDARIA. Madrid: en prensa.

Jornet, J. y Suárez, J. (Coords.) (1997). Evaluación de la Formación Profesional Ocupacional y Continua en la Comunidad Valenciana. Symposium presentado en el *VIII Congreso Nacional de Modelos de Investigación Educativa.* Sevilla: AIDIPE.

Jornet J.M. y Suárez J.M. (Coords.) (1999). *Informe de Evaluación del Proyecto Telematics Training (TETRA).* Valencia: ADEIT / Studio 2000.

Jornet J.M. y Suárez J.M. (2000). Evaluación de Programas de Formación para el Trabajo. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas.* Valencia, Marzo (en prensa).

Jornet, J.; Suárez, J. y Belloch, C. (1998). *Metodología de Evaluación del Programas de Formación Profesional, Ocupacional y Continua.* Proyecto Eurodialogue Evaluation. Valencia: ADEIT.

Jornet, J.; Suárez, J; Perales M.J. y Belloch, C. (1998). Informe evaluativo de los recursos de Formación Continua en el País Valenciano. Valencia: FOREM-PV.

Keep, E. y Mayhen K. (1999). Evaluating the assumptions that underlie training policy. En: J. Ahier y G. Esland (Eds.): *Education, Training and the future of work I.* (págs. 113-140). Londres: Rowtledge.

Marí, R.; Perales, M.J. y Orellana, N. (1997). Orientación y empleo. Análisis del tipo y modelo de las tareas de orientación en los cursos de Formación Ocupacional. En AEOP (comp.) *La Orientación Educativa y la intervención pedagógica integradas en el curriculum.* Valencia: AEOP.

Marí, R.; Perales, M.J. y Villanueva P. (1999). Funciones de la Orientación Ocupacional en la Formación Profesional Ocupacional. Una propuesta. En AIDIPE (comp.) *Nuevas realidades educativas. Nuevas necesidades metodológicas.* Málaga: CEDMA.

Martínez Piñeiro, E. (1995). Problemática actual de la Evaluación de Programas de Formación Ocupacional. Propuesta para su mejora. En AIDIPE (comp.) *VII Seminario de Modelos de Investigación Educativa. Estudios de Investigación Educativa en Intervención Psicopedagógica*. Valencia: AIDIPE.

Mateo, J. (2000). La evaluación de instituciones educativas. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas*. Valencia, Marzo (en prensa).

OCDE (1991). Evaluating Labour Market and Social Programs. París: OCDE.

Perales, M.J. (2000). *Enfoques de Evaluación de la Formación Ocupacional y Continua. Estudio de Validación de un modelo*. Tesis Doctoral. Universitat de València. (en prensa, Microfichas).

Pérez Carbonell, A. (1999). Metodología de la Evaluación de Programas: Evaluación de Programas de Postgrado en la Universitat de Valencia. En CIDE: *Premios Nacionales de Investigación Educativa, 1998*. Madrid: MEC.

Pérez Juste, R. (2000). La evaluación de programas en educación. *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas*. Valencia, Marzo (en prensa).

Tejada, J. (1992). *La evaluación en la Formación Ocupacional. En A. Ferrández (Dir.) La Formación Ocupacional. Realidad y Perspectivas*. Madrid: Diagrama.

Villanueva, P. y Catalá, J.J. (1997). Formación Ocupacional. Aspectos diferenciales y politicas de empleo. En AIDIPE (comp.): *Actas del VIII Congreso Nacional de Modelos de Investigación Educativa*. Sevilla: AIDIPE.

522

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 539-552

EVALUACIÓN DE PROGRAMAS PARA LA EDUCACIÓN DE SOBREDOTADOS

José Cajide Val
Universidad de Santiago de Compostela

1. INTRODUCCIÓN

La evaluación de un programa educativo para niños sobredotados no es una tarea fácil y menos sin contextualizarla en una situación concreta. Por ello a lo largo de este trabajo analizaremos las partes de un programa, las dificultades que presenta y, terminaremos con una propuesta de evaluación siendo conscientes de que va a estar condicionada por múltiples factores: disponibilidades del centro, formación y dedicación de los profesores, características de los alumnos, medios, recursos, etc.

La *evaluación de un programa* para algunos autores consiste en valorar el grado en que el programa está funcionando como se planificó, de tal manera que pueda mejorarse y asegurar su rendimiento de cuentas «acountability», Borland, (1989, p. 211). También puede entenderse como un proceso de análisis y control diseñado para determinar la relevancia, efectividad, significación e impacto de actividades específicas —en este caso para *niños sobredotados*— y el grado de eficacia con que se desarrollan, Doval y Cajide (1995, p. 207). Pérez Juste (2000, p. 554) afirma que la evaluación de programas es un proceso sistemático, diseñado intencional y técnicamente, de recogida de información rigurosa —valiosa, válida y fiable— orientada a valorar la calidad y los logros de un programa como base para la posterior toma de decisiones de mejora, tanto del programa como del personal implicado y, de modo indirecto, del cuerpo social en que se encuentra inmerso.

Asimismo, Pérez Juste dice que en educación un programa encarna un «plan de acción» al servicio de los objetivos valiosos en términos educativos. Por ello, nosotros creemos que dado que la legislación en España acerca de la sobredotación apunta más a la ayuda y mejora del alumno que al rendimiento de cuentas, —tal como veremos posteriormente al referirnos a la evaluación— debemos centrarnos más en la totalidad

del programa, particularmente, en aquellas partes que pueden ayudar a atender más adecuadamente las características propias del alumno sobredotado y, determinados aspectos del mismo considerados relevantes para la evaluación.

2. QUÉ SE ENTIENDE POR SOBREDOTADOS

Si recurrimos a la legislación, en la Orden de 28 de octubre de 1996, DOG, de 28 de noviembre de 1996, se puede leer que «sobredotados» son los que demuestran poseer una inteligencia muy superior a la media y un alto nivel de creatividad. A nuestro entender esta definición encierra bastante indeterminación, al no fijar un nivel intelectual a partir del cual un alumno es sobredotado; y además, el concepto de creatividad es lo suficientemente flexible para ser susceptible de muchos significados distintos. El incluir en la definición una habilidad intelectual general superior, las habilidades intelectuales específicas, el pensamiento creativo y los talentos creativos específicos son bastante habituales en la investigación al respecto, y así lo recoge Milgran (1999, p. 11-12), pero no vamos a entrar ahora en esta controversia. El hecho es que el programa supone una identificación previa adecuada y consensuada de estos sujetos para poder ponerlo en marcha y también para evaluarlo.

3. LA LEGISLACIÓN QUE REGULA UN PROGRAMA DE SOBREDOTADOS

La respuesta educativa dada para los sobredotados por nuestra legislación: a) aconseja una *anticipación* o inicio de la escolaridad obligatoria o una *reducción* del período de escolarización cuando, realizada la evaluación psicopedagógica, se compruebe que el alumno tiene alcanzados los objetivos de etapa, ciclo o curso y, se prevea que esta medida es la más adecuada para su desenvolvimiento personal y social, b) aconseja una adaptación curricular de ampliación dentro del curso que le corresponda por edad cuando realizada la evaluación psicopedagógica, se compruebe que el alumno obtiene un rendimiento excepcional en un número limitado de áreas o cuando, aún obteniendo un rendimiento global alto y continuado se detectaran desequilibrios en el ámbito afectivo y de inserción social.

El procedimiento de cómo hacerlo operativo —siguiendo la legislación— consiste en reducir un año la escolarización en educación primaria (esta no puede tener lugar si ya se anticipó el ingreso en la escolarización obligatoria), y también podrá reducirse un año la escolarización en la educación secundaria obligatoria. La adaptación curricular de ampliación introduce modificaciones significativas en los elementos prescriptivos del currículum común —objetivos, contenidos, y criterios de evaluación— y pueden afectar a una, a varias, o a todas las áreas.

Ahora bien, un programa acorde con la anterior legislación, conlleva una construcción conceptual, una organización que comprende el diseño racional de un conjunto de operaciones y actividades para conseguir resultados con el sujeto/os a que se aplica. Así debemos contar, al menos implícitamente, con un modelo de programa en sentido conceptual, para que exista virtualmente un programa para evaluar y determinar qué efectos se pueden esperar y en qué circunstancias se pueden generalizar.

Entre las ventajas de diseñar programas apoyados en una teoría están: ayudar a comprender conceptual y metodológicamente su ámbito de funcionamiento; servir para definir y esclarecer la forma de identificar los efectos no establecidos en las metas del mismo; ayudar a establecer el perfil y la competencia de las personas que deben participar y, anticipar posibles comparaciones, contrastes y relaciones que deben analizarse.

4. PARTES DE UN PROGRAMA

Es evidente que para poder referirnos a la evaluación de un programa hemos de considerar algunos apartados de éste tales como: 1. Filosofía o teoría en que se basa, 2. Procedimientos y criterios de identificación, 3. Metas y objetivos, 4. Currículo, 5. Recursos que soportan cada una de sus partes, 6. Presupuesto, 7. Dirección, 8. Estrategias instructivas, 9. Evaluación del programa.

4.1. Metas y objetivos

Aparte de la «teoría» que debe fundamentar un programa para superdotados que en nuestro caso viene marcada en parte por la legislación, debemos resaltar que conviene que esté escrito, consensuado, sea viable en sus metas y que refleje los resultados deseados en respuesta a las necesidades del estudiante. El programa debe tener múltiples metas y objetivos flexibles en el tiempo, más que fijos. Para la selección de objetivos y contenidos de adaptación curricular individualizada se deberán priorizar los de mayor carácter funcional e instrumental, los más adaptados a los intereses, capacidades y necesidades del alumno, los objetivos generales básicos de etapa, ciclo o curso, introduciendo otros que amplíen la competencia curricular. En todo caso, las adaptaciones curriculares que se realicen promoverán el desarrollo pleno y equilibrado de las capacidades establecidas en los objetivos generales de la educación obligatoria, (véanse los objetivos generales en la LOGSE, Art. 19), la selección de objetivos de evaluación es un esfuerzo colaborativo entre el equipo de evaluación y los que toman decisiones, las prioridades están determinadas en relación a los propósitos de la evaluación, el tiempo y el presupuesto.

4.2. Currículo en sentido estricto

Es la parte central de un programa, y viene dado por los contenidos curriculares que se pueden ampliar, enriquecer o impartir de forma acelerada elaborando un currículum diferenciado. La diferenciación en los procesos debe hacer hincapié en la flexibilidad de pensamiento, el descubrimiento, la creatividad, la indagación, formulación de preguntas, prueba de hipótesis, y la creación de un ambiente que facilite, la solución de problemas, la divergencia y la flexibilidad en el que se acepten las diferencias individuales, el trabajo personal y cooperativo. Asimismo, se insistirá en el desarrollo de un autoconcepto positivo, la independencia de juicio, los valores sociales, la toma de decisiones, la cooperación mediante el trabajo en equipo... en definitiva, se deben cuidar los ámbitos intelectual y afectivo.

525

Ahora bien han de tomarse algunas decisiones: a) la elección del contenido, ya sea adaptándolo de los libros de texto o bien elaborándolo completamente de acuerdo con las pretensiones del programa, b) la confección de documentos-guía que habitualmente se necesitan para los profesores y la elaboración de materiales para los alumnos y, c) la conveniencia de que los contenidos no interrumpan la marcha normal de la clase ordinaria. Por otra parte, han de evitarse los fallos de diseñar un currículo ceñido exclusivamente al aspecto instructivo y de aprendizaje olvidándose de los sentimientos, la motivación, el concepto de sí-mismo, etc.

4.3. El personal

La aplicación de un programa requiere la identificación previa del alumno/s, profesor/es, ordinario/os y otros expertos que vayan a participar en él. Es necesario que todos los participantes tengan una idea global del programa. La formación del personal que va a impartir cualquier programa para *sobredotados*, en el aula o fuera de ella, es condición *sine qua non* para su implantación. El profesor debe tener autoconfianza, flexibilidad, capacidad para apreciar la valía del sobredotado, cuestionar, estimular, orientar, desarrollar un currículo de calidad, y saber diseñar el aprendizaje. Los padres han de saber respetar las capacidades del niño, tener buena comunicación con el, proporcionarle libertad e independencia para actuar y, los compañeros han de aprender a respetar las diferencias individuales.

También conviene resaltar que todo programa educativo necesita ser «*tutorizado*», asesorado, durante todo su proceso. La tarea de ayuda, de apoyo al mismo, permite durante su desarrollo introducir cambios, distinguir entre factores internos y externos que pueden impedir la implantación. A veces, la cantidad y la calidad de las entradas no son las causas de que un programa fracase sino los factores políticos, sociales, culturales, económicos, el cansancio y otras dificultades. Pero para asegurar la implantación es preciso proporcionar un *feedback* empírico al personal del programa, dado que éste está constantemente afectado por el ambiente en que funciona; por ello, se necesita un «tutor» una persona que crea realmente en él e irradie ese interés en los demás implicados.

4.4. El presupuesto

No basta con la disposición favorable del profesorado hacia el programa, la puesta en marcha genera gastos y, aparte de la inversión en la formación específica del personal del programa, se necesita la disponibilidad de un determinado material complementario del que debe hacerse previsión en el centro y que es preciso tener en consideración si queremos que el programa funcione.

4.5. La evaluación

La legislación a este respecto afirma que la evaluación de un programa que siga una adaptación curricular se hará en función de los criterios de evaluación individua-

lizados establecidos en ella, sin perjuicio de que para los efectos de promoción el referente sean siempre los objetivos establecidos para la etapa correspondiente, Orden 8 de octubre de 1995 (DOG de 7 de noviembre).

La evaluación de los aprendizajes de este alumnado, en aquellas áreas o materias que hubieran sido objeto de adaptaciones curriculares significativas se efectuará tomando como referencia los objetivos y criterios de evaluación fijados para ellos en las adaptaciones correspondientes, conforme establece en el apartado segundo, la Orden de 14 de febrero de 1996 sobre la evaluación de los alumnos con necesidades educativas especiales que cursan las enseñanzas de régimen general establecidas en la Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo. Resolución de 29 de abril de 1996 (BOE de 16 de Mayo de 1996).

Por tanto, los criterios de evaluación individualizados establecidos en la adaptación curricular y los objetivos de la etapa deben servir de pauta en la evaluación del programa.

4.5.1. La dificultad de la evaluación de un programa para sobredotados

Para facilitar la evaluación ha de existir acuerdo respecto: a) el concepto de sobredotado por tener diferentes significados e interpretaciones, b) precisión respecto a el tipo de resultados del programa que deberían alcanzarse, c) tomar decisiones respecto a los instrumentos estandarizados o que es preciso construir para evaluar, y d) el establecer elementos de comparación. Si estos factores no reciben la adecuada atención es difícil determinar cuando podemos considerar que un programa tuvo éxito.

Cuando nos enfrentamos a un programa individualizado, donde las metas varían de individuo a individuo y, las estrategias son adaptadas al alumno, su evaluación es todavía más difícil y para, facilitarla debe existir consenso acerca de lo que es un niño sobredotado de lo que debe hacerse con el, y de cuáles son los resultados esperados.

4.5.2. Factores que condicionan la evaluación

Primero, los destinatarios de evaluación, especialmente los que toman decisiones y aquellos a quien sirve el programa si esperamos que la evaluación tenga algún impacto.

Segundo, la valoración continua de la consecución de los objetivos curriculares amplios que habitualmente pretenden incrementar la creatividad, abordar la solución de problemas, desarrollar el pensamiento crítico, la independencia, la flexibilidad...; así como de los específicos que reflejan el contenido del currículo, es decir, están más ajustados a los objetivos de etapa, curso, etc. Una evaluación de sobredotados debe incluir ambos.

Tercero, los encargados de poner en marcha un pograma necesitan cierto ingenio para desarrollar objetivos amplios para el estudiante dado que los habituales de la clase regular no deben presentar problema para ellos. La ayuda de «personas interesadas» en el programa: padres de los bien dotados, estudiantes, profesores y directivos pueden contribuir a clarificarlos e identificarlos cuando se pretende tener referencias

527

generales acerca de la efectividad del programa. Pero cuando los propósitos son más concretos, si los objetivos del programa están definidos y los que toman decisiones quieren conocer si se han conseguido, la evaluación deberá dar cuenta de ellos.

Cuarto, el equipo de evaluación deberá elegir los criterios que son medidas razonables y prácticas de los elementos de evaluación. Si los criterios son orientados a resultados, la evaluación deberá ser más empírica, usando un enfoque de evaluación que requiere reunión de datos y análisis, más de tipo experimental, o cuasi-experimental. Si los criterios son más orientados a procesos el investigador usará un enfoque más cualitativo, por ej., evaluar los proyectos realizados, documentos escritos, observaciones de clase..., que existen acerca del programa, analizarlos y compararlos contra ciertos estándares predeterminados. Las técnicas cualitativas deben incorporarse aceptando que los evaluadores están obligados a demostrar que «las valoraciones del desarrollo» se realizan con criterios válidos tales como evidencia y concordancia con otras evaluaciones, deben emplearse diferentes enfoques pero bajo criterios de validez, fiabilidad y comparación (Callahan, 1993, pp. 614 y 615).

Quinto, también las medidas de habilidades de pensamiento crítico, creatividad..., son asimismo una parte de la evaluación de un programa tal como se deriva de la propia definición de superdotado y de la legislación. Más difícil es conseguir medidas válidas y fiables que evalúen el currículo en su totalidad, pero aún siendo una tarea laboriosa la construcción de instrumentos es una tarea necesaria para la evaluación.

4.5.3. Algunas consideraciones respecto a la evaluación de programas

La reunión de datos y su evaluación depende de que las distintas partes de un programa estén claramente definidas. Considerar si el propósito de la evaluación es formativa, destinada a mejorar el programa, o sumativa si se centra específicamente en los resultados sin más, Gowan (1999, pp. 1-5), afecta al diseño de evaluación.

Existen algunos peligros que debemos considerar para evaluar los efectos de un programa, a veces no evaluamos tanto el programa en sí cuanto al profesor o un pequeño grupo de profesores, que intervienen. También las aptitudes interactúan con el tratamiento, a veces la misma instrucción produce efectos distintos dependiendo del tipo de estudiante.

Las potenciales soluciones para evitar los peligros citados consisten en extremar el control y combinar estrategias cuantitativas y cualitativas. Las estrategias de evaluación deben proporcionar la oportunidad de describir el camino en que los sobredotados con ciertas características interactúan con las experiencias previstas para ellos. Los estudios cualitativos son necesarios para determinar qué tipo de trabajos, para qué individuo, bajo qué condiciones intervienen ayudando al alumno sobredotado según sus necesidades.

5. EL DISEÑO DE LA EVALUACIÓN

El diseño es una estrategia de investigación destinado a tratar de conocer cuál es la efectividad de un programa y si es este programa mejor que aquel otro para sobre-

dotados. Constestar estas preguntas tiene que ver con el diseño de evaluación y la forma como reunimos la información. Cuestionarse si el programa para sobredotados es mejor que seguir la marcha normal de clase no tiene mucho sentido porque los bien dotados deben estar atendidos en los centros en cualquier situación, incluso si el programa para sobredotados no existe. Ahora bien, no es fácil probar que algo es efectivo si no existe alguna comparación. Así comparamos un programa con otro, o con una cantidad de estándares (o criterios), o bien el grado en que el programa responde a las metas desarrolladas por los educadores y profesores para ese programa.

Es posible pensar en varios tipos de diseños pero posiblemente los cuasi-experimentales en su formato de diseños de series de tiempo son los que mejor se ajustan a la filosofía defendida por nuestra legislación.

5.1. Diseños experimentales

Los diseños experimentales al azar con grupo de control pretest-postest (véase cuadro 1, apartados 1.1); y si no existe pretest (véase apartado 1.2), representan esquemas de investigación que aplicados a sobredotados presentan algunos problemas. En efecto, las situaciones educativas hacen difícil conseguir dos grupos al azar de niños sobredotados, asimismo parece poco procedente que un grupo experimente el programa frente a otro que no lo hace (en realidad deberán experimentar el programa de la clase regular, o bien uno equivalente puesto que no se puede dejar sin programa). Las ganancias finales si existen posiblemente no hagan sino confirmar las diferencias iniciales previas. Cabe trabajar con dos programas equivalentes y ver las ganancias de cada grupo respecto a la situación inicial tratando de indagar a quien favorece realmente el programa y extraer enseñanzas del análisis de proceso, además del estudio de los resultados.

CUADRO 1
DISEÑOS APLICABLES A LA EVALUACIÓN DE SUPERDOTADOS

1. **Diseños experimentales:**

1.1. Diseño de grupos al azar con control pretest-postest.

A	O_1	X	O_2
A	O_3		O_4

1.2. Diseño de grupo de control con postest solamente.

A	X	O_1
A		O_2

529

2. Diseño contrabalanceado.

Grupos A y B (sobredotados). El B sirve de comparación.

Grupo A			Grupo B		
Tiempo 1	Tiempo 2		Tiempo 1	Tiempo 2.	
S_1	X_1	Y_1	S_4	Y_1	X_1
S_2	X_2	Y_2	S_5	Y_2	X_2
S_3	X_3	Y_3	S_6	Y_3	X_3

Grupos C (no sobredotados). Grupo D (no sobredotados) El D sirve de comparación.

T_1	X_1	Y_1	T_4	Y_1	X_1
T_2	X_2	Y_2	T_5	Y_2	X_2
T^3	X_3	Y_3	T_6	Y_3	X_3

3. Diseños cuasi-experimentales:

3.1. Diseños de series temporales.

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

3.2. Diseños de línea base múltiple.

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8
O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8
O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

3.3. Diseños de grupo de control no equivalente.

O_1 X O_2
O_3 X O_4

4. Diseño correlacional.

O_1 O_2

5.2. Diseños contrabalenceados

Los estudiantes de un programa (véase cuadro 1 apartado 2) de sobredotados pueden repartirse en grupos pequeños y el currículo también se divide en unidades que se aplican en diferentes tiempos a lo largo del año. El procedimiento es como sigue: el grupo **A** recibe un currículo **X** para niños sobredotados mientras el grupo **B** (también de sobredotados) recibe otro currículo **Y**. La evaluación tiene lugar al final del tiempo marcado. Así el grupo que recibe el tratamiento (currículo) **Y**, actúa de control para la

530

efectividad del currículo **X** y, el grupo **X** actúa de control para el currículo **Y**. Pero no quedaría demostrada la efectividad de un currículo diferenciado; por ello se añaden los grupos **C** y **D** de alumnos no sobredotados, que reciben el programa de los sobredotados de maestros también de sobredotados. Así el grupo **C** sirve de comparación para el grupo **A**, y el grupo **D** de comparación del grupo **B**. Esto permite al evaluador comparar el desarrollo de los estudiantes de clase regular con el de los estudiantes sobredotados y, asimismo comprobar si los alumnos de clase regular pueden aprovecharse del programa de los sobredotados en la misma medida que estos, (véase Carter, 1991, p. 244).

5.3. Diseños de series de tiempo

Estos diseños son esquemas de investigación susceptibles de adaptaciones teniendo en cuenta la realidad escolar. Por ejemplo, hacer medición (evaluación previa), aplicación del programa y medida durante la aplicación, es decir, evaluación del proceso y medición una vez finalizado, de la eficacia en distintos niveles, esto se puede hacer semanal, mensual, trimestral y anualmente; de este modo podemos comprobar cual es la evolución del sujeto a lo largo del programa, analizar en que medida se han conseguido unos objetivos, y comparar las medidas de distintas etapas del proceso, etc.

En estos diseños aplicables a un sujeto único o bien a un grupo, las medidas se representan por la letra **O** mientras que para intervención se recurre a la letra **X** (ver cuadro 1, apartado 3). Presentan algunas ventajas, a), permiten evaluar una o varias conductas, dado que pueden tener más de una línea base simultáneamente, b) son susceptibles de análisis cuantitativos y cualitativos, c) permiten establecer comparaciones entre conductas del propio sujeto, o con conductas de otros sujetos, y d) las propias medidas del sujeto a lo largo del tiempo sirven como su propio control.

Los diseños de línea base múltiple, adaptados, son susceptibles de amoldarse para introducir varios tratamientos a lo largo de un proceso (estos pueden ser las distintas unidades didácticas), asimismo valorar el efecto sobre distintas conductas de un mismo sujeto, pudiendo apreciar su evolución. En teoría podrían incluso ser conductas de sujetos diferentes que luego podemos comparar.

El análisis puede ser cuantitativo y cualitativo; si las comparaciones se hacen entre fases de tratamiento y no tratamiento recurrimos a la prueba de «t»; si tenemos medidas de antes, durante y después del tratamiento a! análisis de varianza y, a la regresión cuando en base a medidas previas hacemos una predicción de los resultados futuros y comparamos ambos. Las limitaciones al uso de estas pruebas está en la falta de independencia de las puntuaciones, dado que al ser medidas sucesivas de los mismos sujetos tienen efectos aditivos que se confunden con el efecto del programa. De ahí que proceda hacer en paralelo tanto estudios más cualitativos, como análisis de tareas, materiales elaborados, gráficos evolutivos de la conducta, etc., que puedan responder a una adecuada evaluación del programa.

531

Otra de sus limitaciones está en que los padres no quieren que se interrumpa un programa y han de llevarse a cabo adaptaciones para conseguirlo.

5.4. Diseños de grupos de control no equivalente

Aquí nos encontramos con un diseño de grupo donde los procedimientos de azar no son posibles. En este caso se toman dos grupos intactos que se prueban antes y después del tratamiento. Mientras un grupo experimenta el tratamiento el otro actúa como control (se supone que siguiendo el de la clase regular o uno alternativo). La interpretación de los efectos del tratamiento requiere comprobar las diferencias previas que existen en variables como edad, habilidad, motivación. Así, el análisis se presenta más fácil si los grupos muestran una equivalencia previa, pudiendo compararse en los resultados finales mediante una prueba de «t». En caso de que no se conozca dicha equivalencia previa conviene utilizar el análisis de covarianza para estudiar las medidas de postest teniendo en consideración las diferencias iniciales.

5.5. Diseño correlacional

Es un diseño de un solo grupo y los datos son analizados vía estadística correlacional; correlación múltiple. Es más apropiado para el estudio de relaciones pero menos indicado para el estudio de los efectos de un programa.

6. PROPUESTA DE EVALUACIÓN DE UN PROGRAMA EN EL ÁREA DE TECNOLOGÍA

A continuación presentamos algunas pautas para la elaboración de un programa y su propuesta de evaluación en el Área de Tecnología. Se trata de una propuesta incompleta, pues un proyecto acabado rebasa los límites de este trabajo. Pese a ello, consideramos que en él se concreta parte de la filosofía del tipo de evaluación de programa defendido a través de las páginas previas. Figura el *proyecto didáctico* que debe completarse en muchas de sus partes, el *diseño de la unidad didáctica* en el que se plantea la solución de un problema/s, los *objetivos*, los *bloques de contenido* y, la *evaluación*: con referencias a su concreción en los instrumentos, criterios, desenvolvimiento instructivo, logros, y valoración global, (ver los Cuadros 2, 3 y 4).

Se presentan asimismo los objetivos, amplios y específicos, se intenta un desarrollo pleno y equilibrado de distintas capacidades y habilidades. Se hace hincapié en el descubrimiento, la flexibilidad, la indagación, la divergencia, pero también en la solución de problemas concretos, se dan pautas para la evaluación del proceso pero también de los logros. Se da importancia al trabajo personal y en colaboración, al uso de la tecnología, a la conexión de los conocimientos teóricos con la vida práctica y, a la solución de problemas.

CUADRO 2
UNA PROPUESTA DE PROGRAMA Y SU EVALUACIÓN

En el Diseño Curricular Base para la Educación Secundaria Obligatoria la Tecnología se define como un conjunto de procesos de invención, fabricación y uso de objetos que se crean o utilizan a la hora de resolver un problema. En esta área su metodología se centra en el método de proyectos.

1. Proyecto didáctico.
Identificación y diagnóstico de necesidades del alumno sobredotado.
Objetivos:
- Generales: planificar un proyecto, desarrollar actitudes de investigación.
- Específicos.
- Conductuales.
- Procedimentales.
- Diseño de actividades.

Organización, metodología y temporalización.
Recursos.
Evaluación.

2. Diseño de la Unidad Didáctica.
- Planteamiento del problema, (elaborar la maqueta de un edificio/s con la ayuda de un programa informático).
- Búsqueda de la información.
- Diseño (reunir las características que deben tener los proyectos).
- Construcción del objeto tecnológico diseñado.
- Memoria de evaluación del proceso seguido para alcanzar la solución.

3. Objetivos de la Tecnología de la ESO: desenvolver capacidades para:
- Abordar de forma autónoma los problemas.
- Utilizar sus conocimientos sobre técnicas básicas.
- Planificar y llevar a cabo proyectos.
- Expresar y comunicar las soluciones.
- Mantener una actitud de indagación, curiosidad, flexibilidad e innovación.
- Valorar la importancia del trabajo en equipo.

4. Bloques de contenidos:
- Resolución técnica de problemas.
- Análisis de objetos y sistemas.
- Representación gráfica.
- Herramientas y técnicas de construcción.
- Técnicas de administración y gestión.

5. Temporalización.
- Materia anual.
- Registro de observación quincenales y final

6. Evaluación.
- Valorar el grado de conocimiento e interés con que se desarrolló.
- Constatar si es capaz de abordar con autonomía las tareas de diseño de la solución a un problema.
- Comprobar si se alcanza el nivel de identificación de formas, dimensiones, materiales empleados.
- Comprobar la capacidad de expresión por medios gráficos.
- Valorar la capacidad de planificación, construcción aplicación, colaboración.

7. Instrumentos de evaluación:
- Observación directa en la clase.
- Observación del proceso: ajuste entre el plan y el desarrollo, resultados parciales, satisfacción.
- Observación del producto final: funcionamiento, creatividad, nivel alcanzado.
- Memoria informe:
  presentación, uso del lenguaje, razonamiento, organización, originalidad.

8. Criterios de valoración:
- ***Excelencia*: superioridad respecto a los demás en alguna dimensión.**
- ***Originalidad*: respuestas no habituales «únicas».**
- ***Productividad*: contrastada o potencial.**
- ***Demostrabilidad*: mediante pruebas válidas y fiables.**
- ***Valor*: mostrar un desarrollo superior validado por la sociedad.**

533

CUADRO 3
EVALUACIÓN DEL DESENVOLVIMIENTO INSTRUCTIVO DEL ALUMNO

Evaluación del proceso instructivo del alumno (Área de tecnología).	
1. Forma de trabajar. — Trabaja cooperativamente. — Pregunta al compañero. — Pregunta al profesor — Sigue al compañero. — Lidera a los compañeros. — Trabaja solo. **2. Relación con los compañeros.** — Positiva. — Neutra. — Negativa. **3. Aportación crítica.** — Colabora constructivamente. — Adopta posturas neutrales. — Interviene destructivamente. **4. Interacción con el profesor:** — Independencia. — Dependencia. — Solicita ayuda. — Aporta soluciones. **5. Interacción con los compañeros.** — Independencia. — Dependencia. — Solicita ayuda. — Aporta soluciones.	**6. Actitudes y forma de trabajo:** — Perseverancia — Participación. — Interés. — Disponibilidad. — Concentración. — Iniciativa. — Creatividad. — Flexibilidad. **7. Presentación de los logros fijados.** — Elaboración del proyecto. — Representación de figuras. — Elaboración de presupuestos. — Presentación de un presupuesto en una hoja de cálculo. — Presentación del Informe de un proyecto con un tratamiento de texto. **8. Hábitos y ritmos de trabajo.** — Curiosidad. — Originalidad. — Colaboración con los demás. — Concentración. — Dispersión. — Ritmo de trabajo lento. — Ritmo de trabajo rápido. — Autonomía. — Independencia.

CUADRO 4
VALORACIÓN GLOBAL EN EL PROGRAMA DE LAS HABILIDADES PARA DISEÑAR, DESARROLLAR Y EJECUTAR UN PROYECTO

Actividad	**Resultados**	**Puntos de observación**	**Criterios: demostrabilidad**	**Proceso de desarrollo: productividad**	**Análisis**
Desarrollar un proyecto (maqueta de un edificio) (Tecnología)	Presentación de un presupuesto — Informe de investigación.	Final de la unidad.	Valoración por: — profesores. — compañeros. — expertos.	— Completa parcialmente. — Completa totalmente.	— Satisfactorio. — Excelente. — Original. — Valioso.

Así la evaluación de un programa de sobredotado/s, debe ser: *multidimensional* valorando habilidades de planteamiento de problemas, búsqueda, diseño, construcción, solución e innovación y evaluando las habilidades tal y como se emplean en el contexto de su aprendizaje; *longitudinal* buscando en el proceso el modo como va desarrollando su entendimiento y su habilidad para ampliar su pensamiento por conexión a métodos, fuentes etc. y; *crítica* que proporcione una respuesta al estudiante ofreciéndole una perspectiva analítica de su trabajo y la posibilidad de mejorarlo de forma innovadora y creadora.

BIBLIOGRAFÍA

Borland, J.H. (1989). *Planning and implementing program for the gifted.* Teachers College, Columbia University, New York.

Buchanan , N.K. & Feldhusen J.J. (1991), (Eds.). *Conducting research and evaluation in gifted education.* Teachers College, Columbia University, New York.

Callahan, C.M. (1993). Evaluation programs and procedures for gifted education: international problems and solutions. In K.A. Heller; F.J. Möns and A.H. Passow. (Eds.). *International handbook of research and development of giftedness and talent.* Pergamon, Oxford.

Carter, K.R. (1991). Evaluation of gifted programs. En Buchanan, N.K. & Feldhusen, J.J. (1991). (Eds.). *Conducting research and evaluation in gifted education.* Teachers College, Columbia University, New York, pp. 245-274.

Cajide Val, J. (1983). *Medida de la creatividad: genios superdotados y creativos.* Servicio de Publicaciones. Universidad de Santiago.

Doval, L. y Cajide, J. (1995). Superdotación: evaluación o diagnóstico. *Revista de Ciencias de la Educación.* 162, abril-Junio, pp. 191-213.

Gowan, J. (1999). *A Handbook of techniques for formative evaluation.* Kogan Page Limited. London.

LOGSE, (1990), Art. 3, 5, 36, y 37.

Louis, B; Lewis, M; Subotnik, R; & Breland, P. (1999). Establishing criteria for high ability vs. selective admission to gifted programs: implications for policy and practice. Paper presented 2000 Annual Meeting. American Educational Research Association. New Orleans, Louisiana, April 24-28.

Milgran, R.M. (1999). (Ed.). *Counselling gifted and talented children.* Ablex Publishing Corporation Norwood, New Jersey.

Orden de 24 de Abril de 1996, BOE de 3 de mayo de 1996.

Orden de 6 de Octubre de 1995, DOG de 7 de Noviembre de 1995.

Orden de 28 de Octubre de 1996, DOG de 28 de Noviembre de 1996.

Orden de 24 de Julio de 1998, DOG de 31 de Julio de 1998.

Orden de 14 de Febrero de 1996, BOE de 23 de Febrero de 1996.

Orden de 2 de Abril de 1997, BOE de 4 de Abril de 1997.

Orden de 24 de Abril de 1996, BOE, de 3 de Mayo de 1996.

ORR, C. (1994). Table 12 Using Achievement Test Scores for judge the efficacy of gifted programs: caution, caution, caution. *Annual Meeting. AERA,* New Orleans, April 4-8, 1994.

535

Pérez Juste, R. (2000). La evaluación como medio para la mejora de la eficacia y de la calidad del aprendizaje, de la educación y de las instituciones. En XII Congreso Nacional y I Iberoamericano de Pedagogía. Tomo 1 Ponencias. Madrid. Pp. 535-572.

Real Decreto 696/1995, BOE de 2 de Junio de 1995.

Resolución de 29 de Abril de 1996, BOE de 16 de mayo de 1996.

Ruestow, M.E. (1994). Table 11 Wadleigh scholarship program. *Annual Meeting. AERA*, New Orleans, April 4-8, 1994.

Subotnik, R. (1994), (Ed.). *Beyond Terman, contemporary longitudinal studies of giftedness and talented*. Ablex Publishing Corporation Norwood, New Jersey.

Witham, J.H. (1994). Table 10 Acceleration: does it happen more frequently for gifted students in private or public schools? *Annual Meeting. AERA*, New Orleans, April 4-8, 1994.

536

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 553-563

EVALUACIÓN DE PROGRAMAS PARA ALUMNOS SUPERDOTADOS

Carmen Jiménez Fernández
Catedrática. UNED
Con la colaboración de Leonor Prieto Navarro y
Marta Sandoval Mena. Universidad Pontificia Comillas. Madrid

RESUMEN

La atención a los alumnos más capaces está conociendo un renovado interés en el marco de una educación para la excelencia que contempla formalmente la diversidad del alumnado. La investigación más rigurosa debe apoyar y modular su educación.

Este artículo resume resultados de la evaluación de programas para los alumnos superdotados, deteniéndose en dos estudios señeros objeto de evaluación longitudinal: el ya histórico de Terman y el modelo del talento matemático de Stanley. Asimismo analiza los resultados del programa de enriquecimiento de Renzulli y los que se obtienen con la agrupación por capacidad en sus diversas manifestaciones.

ABSTRACT

Educational attention of gifted students is achieving renewed interest within the frame of education for excellence, since it contemplates student diversity formally. Rigorous research must support and modulate the education of such students.

This study summarises the evaluation results of programs for the gifted, placing special emphasis on two crucial longitudinal evaluation programs, Terman's already historical and Stanley's model of mathematical talent. It also analyses the results of Renzulli's enrichment program and those obtained through ability grouping in its different forms.

537

1. PLANTEAMIENTO DEL TEMA

Todo programa pretende satisfacer las necesidades de sus destinatarios que en nuestro caso son alumnos que aprenden a mayor ritmo, con mayor profundidad y mayor extensión que sus iguales. La evaluación de programas no es una práctica habitual de la escuela. La alta capacidad afecta a una proporción menor de la población escolar que, además, no goza de pleno reconocimiento formal. Sin embargo, es objeto de programas específicos en países como Estados Unidos, programas que suelen incorporar un dispositivo evaluador que pretende documentar su impacto, justificar su coste, identificar los elementos críticos y ofrecer indicadores de su eficacia a los grupos a los que concierne el programa, cara a tomar decisiones de mejora respecto del programa y por parte de las personas en él implicadas.

En las líneas que siguen se esbozan los aspectos nucleares en la evaluación y se resume la valoración que merecen programas para los alumnos más capaces y que giran en torno a tres ejes clásicos: aceleración, enriquecimiento y agrupamiento por capacidad.

2. JUSTIFICACIÓN DE LA EVALUACIÓN DE PROGRAMAS PARA ALUMNOS SUPERDOTADOS

La evaluación de estos programas entraña dificultades propias. Por ejemplo, los tests para el diagnóstico de la alta capacidad tienen un techo muy bajo; características como las preferencias sobre el estilo de vida que modulan las opciones curriculares de las jóvenes superdotadas apenas si están formalizadas en instrumentos de medida (Lubinski, Benbow y Sander, 1993). En programas de aceleración es fácil evaluar si un alumno de doce años logra los mismos resultados que otro de trece, o si el 90 por 100 de los más capaces ingresa en prestigiosas universidades frente al 30 por 100 de la población general. También es fácil evaluar productos que reflejen con claridad cambios en las habilidades del alumno y que, probablemente, no habrían adquirido de no mediar el programa. Es más difícil evaluar la capacidad de tomar decisiones, el pensamiento crítico, la autoconciencia, la responsabilidad social, la motivación intrínseca, la creatividad, aspectos definidores de varios programas. Estudiosos de la superdotación proponen modelos de evaluación de programas para dichos alumnos, tal es el caso de Callahan (1993, 1998), Van Tassel-Baska (1992) o Davis y Rimm (1998). Callahan y Caldwell (1995) señalan ocho finalidades a legitimar en dicha evaluación que agrupan como sigue:

1. *Planificando para la evaluación*. Prever desde el comienzo información útil para documentar la necesidad del programa, justificar el modelo específico de programa elegido y mostrar que es factible de aplicación.
2. *Evaluación de la puesta en marcha del programa*. Recoger información útil para documentar cómo se ha aplicado de hecho el programa e identificar sus características positivas y sus debilidades cara a mejorar el programa y su aplicación.
3. *Evaluación del programa ya estabilizado*. Mostrar los resultados del esfuerzo seleccionando los aspectos que más conciernen a cada uno de los grupos que han sostenido el programa como profesores, dirección, padres, autoridades y patrocinadores externos.

533

3. EL ENFOQUE LONGITUDINAL EN LA EVALUACIÓN DE PROGRAMAS

Se trata de una familia de métodos de medidas repetidas que recogen datos de los mismos sujetos en las mismas o comparables características, en dos o más periodos temporales distintos. Observaciones múltiples de los mismos sujetos a lo largo del tiempo permiten el análisis de las diferencias intra e interindividuales y determinar la eficacia de un programa (Subotnik y Arnold, 1993). Analicemos dos estudios señeros, si bien el primero no es propiamente un programa sino un marco adecuado para comprender la alta capacidad.

3.1. El estudio longitudinal de Terman

Fue iniciado en 1921 en la Universidad de Stanford en California y su finalidad era analizar las características de los sujetos con alto CI, la evolución de sus logros y el grado de estabilidad de la inteligencia. El procedimiento fue el siguiente. Seleccionó una muestra de 1.528 alumnos con edades comprendidas entre los 12 y los 14 años; 857 alumnos y 671 alumnas. Utilizó dos criterios. En primer lugar pidió al profesorado que preseleccionara a los tres alumnos más dotados. En segundo lugar, aplicó a los preseleccionados el test de Stanford-Binet y seleccionó a los que obtuvieron en él 140 o más puntos. Seleccionó un grupo de control de la población general.

Obtuvo una serie de medidas en distintos momentos a través de cuestionario para los padres, informe de los maestros, examen médico de aproximadamente la mitad de la muestra, baterías de tests de rendimiento y de personalidad, inventario de intereses personales, medida de las actitudes hacia el juego, medidas antropométricas estándar y entrevistas. Logró el seguimiento del 90 por 100 de la muestra inicial a lo largo de medio siglo, pues sus colaboradores publicaron resultados después de fallecido Terman (Oden, 1968).

El resumen (Terman, 1925, Terman y Oden 1947; Oden, 1968) de los obtenidos puede ser que en relación al grupo normal, los superdotados destacaban significativamente en rendimiento académico y profesional, no manifestaban problemas de ajuste personal, destacaban ligeramente en características físicas como salud y aspecto; tenían intereses más elevados, eran mejores lectores y procedían de familias y ambientes más selectos. Son valores de grupo, es decir, había sujetos que en la edad adulta ejercían trabajos intelectualmente sencillos y otros sobrepasados por las circunstancias de la vida pero, en conjunto, ocupaban profesiones en los niveles altos en términos significativamente superiores a la población general y presentaban buena adaptación. Rompió el mito de la fragilidad física del superdotado y estableció con claridad la correlación entre rendimiento escolar y CI.

Algunas precisiones para valorar dichos resultados (Laycock, 1979; Vialle, 1994). Partió de una concepción unidimensional de la inteligencia muy vinculada al rendimiento académico; no tomó en consideración la creatividad y quizá por ello en su estudio no hubo ningún genio, en el sentido de persona tan capaz que rompe con las normas imperantes en un campo y termina por crear las suyas propias; los sujetos se sabían positivamente observados; muestra general y submuestras contenían sesgos

como hiperrepresentación de los judíos e hiporrepresentación del género femenino; no prestó atención a la mortalidad experimental; el grupo de control no era equivalente al experimental[1].

3.2. El talento matemático o modelo de Stanley

El SMPY (*Study of Matematically Precocious Youth*) o estudio del talento matemático ha sido iniciado por Julian C. Stanley en 1971 en la Universidad Johns Hopkins de Baltimore y su finalidad es conducir una investigación que sirva para ayudar a los estudiantes intelectualmente muy capaces en el campo matemático, identificando los factores que contribuyen a su desarrollo académico y vocacional (Stanley, 1973). Es un modelo con relieve propio por la capacidad que estudia, la matemática o científica; por el modelo de identificación y diagnóstico que ha desarrollado, la «búsqueda del talento» mediante tests de nivel superior; por la opción curricular en que se apoya, «la aceleración a la medida», y por incorporar un diseño longitudinal multicohorte que a lo largo de cincuenta años pretende valorar la trayectoria de los participantes y estudiar diferencias intra e interindividuales (Benbow, 1991). En 1972 Stanley y su equipo identificaron 450 estudiantes; veinte años mas tarde la cifra de aspirantes había superado los 140.000.

El proceso de diagnóstico tiene dos fases. En la primera, seleccionan a los alumnos que a la edad de 12-13 años se sitúan en el percentil 97 o superior en los tests estandarizados de rendimiento en los baremos propios de esa edad. Los alumnos del 3% superior pueden ser candidatos al SMPY. En la segunda, se les aplica el SAT-M (*Scholastic Aptitude Tests*) o test de evaluación de la capacidad matemática versión propia de los alumnos de 16-17 años. El grupo del 3% superior es de por sí altamente selecto; sin embargo obtiene en el test de nivel superior o con techo más alto (SAT-M), puntuaciones que varían ampliamente y que se distribuyen a su vez según el modelo de la curva normal (Cfr. Jiménez Fernández, 2000, cap. I y III).

Aunque el programa ha revisado sus planteamientos, permanece el concepto inicial y considera altamente capacitados sobre el 20%-30% de los preseleccionados, que equivale a entre el 1% y el 0,5% de la población general. Los restantes alumnos se categorizan estableciendo puntos de corte en el tests de nivel superior. Tras unos años de aplicación purista, el SMPY se ha abierto a la investigación del talento verbal y selecciona a los aspirantes con la misma estrategia, sólo que les aplica el SAT-V o test de razonamiento verbal de nivel superior (Benbow, 1991; Cohn, 1991). Las puntuaciones del SAT se expresan en una escala que varía de 200 a 800 con media de 500 y desviación típica de 100. El programa continúa y ha sido incorporado por otras instituciones. En 1986 se dividió en dos ramas, creando una similar en la Universidad de Iowa liderada por Benbow.

---

1 Uno de los aspectos criticado *a posteriori* ha sido la interpretación dada a las diferencias entre géneros, totalmente ligada al estereotipo social. En la edad adulta, la mitad de las mujeres superdotadas estudiadas eran amas de casa y la otra mitad realizaba además tareas profesionales en puestos menos remunerados, en trabajos de inferior rango académico o con dedicación parcial. Terman era hijo de su tiempo. Las pruebas de CI no mostraban diferencias significativas entre hombres y mujeres y un dato de marca como este, le hacía dudar de las explicaciones simplistas; pero fue incapaz de avanzar nuevas hipótesis explicativas y se dejó atrapar por sesgos y estereotipos (Cfr. Vialle, 1994).

3.2.1. La aceleración como opción curricular

Stanley considera la aceleración como la mejor opción para los alumnos altamente capaces, coincidiendo en ello con las recomendaciones de Terman[2]. En la práctica, la

CUADRO 1
*TIPOS DE PROGRAMAS RECOMENDADOS EN EL **SMPY** SEGÚN LOS RESULTADOS OBTENIDOS EN EL SAT (COHN, 1991)*

a) Puntuaciones en el límite de las esperadas por azar o de aproximadamente 230 en el SAT-V y de 270 en el SAT-M. Los participantes en el «talent search» o búsqueda del talento que obtienen puntuaciones en el límite de lo esperado por azar, precisan recordar que aún así pertenecen al grupo del 3 % o del 5 % superior de su grupo de edad en las pruebas de razonamiento matemático y verbal, respectivamente. Las opciones educativas apropiadas para estos chicos son:

- Cursar al máximo nivel (*honors-level*) materias fundamentales.
- Seminarios de enriquecimiento, tales como estudios interdisciplinares.
- Oportunidades educativas extraescolares ofrecidas por la comunidad.
- Orientación académica.

b) Puntuaciones entre 270 y 500 en el SAT-M o entre 230 y 430 en el SAT-V. Equivalente en el primer caso a la media obtenida por los alumnos varones de bachillerato-COU y en el segundo a la media obtenida por dichos alumnos de ambos sexos. Recomiendan:

- Comenzar tempranamente a ritmo regular un curso de un tema importante, sobre todo para los que puntúan en el rango más bajo.
- Compactar el currículo en alguna materia importante realizando en un año el equivalente a dos.
- Curso de verano en la universidad o clase a lo largo del año académico, para los muy motivados y situados en el rango superior del intervalo.

c) Puntuaciones entre 510 y 630 en el SAT-M o entre 440 y 590 en el SAT-V. Estos alumnos puntúan por encima de la media de los alumnos de bachillerato-COU. Recomiendan:

- Enseñanza individualizada a ritmo rápido y de contenido complejo en aspectos importantes del currículo.
- Realizar algún curso universitario mientras culminan los estudios de secundaria.
- Acceso precoz a cursos reglados de nivel superior.
- En algunos casos, salto de curso cuidadosamente planificado.

d) Puntuaciones entre 640 y 800 en el SAT-M o entre 600 y 800 en el SAT-V. Son extremadamente dotados en capacidad numérica o verbal y además de las opciones anteriores que son posibles en todos los casos, recomiendan la:

- Admisión temprana a tiempo parcial en la universidad.
- Admisión temprana a tiempo total en la universidad.

2 Su colaboradora Benbow (1991) recoge las siguientes formas de aceleración: salto de curso; entrada precoz en la escuela preescolar o primaria; condensar el currículo y realizar tres cursos en dos años, por ejemplo; currículo de progreso continuo; ingreso precoz en bachillerato o en la universidad; realización de cursos superiores en programas de verano, horario nocturno o similar en paralelo con el bachillerato; cursos reglados de nivel avanzado y enseñanza dual; mentores y cursos a distancia que se prestan bien para enriquecer y acelerar el currículo pues si van preparados, son alumnos con gran capacidad de aprendizaje autónomo.

BEST COPY AVAILABLE

recomendación de un tipo u otro de enseñanza se apoya básicamente en los resultados obtenidos por los alumnos en el SAT. El Cuadro 1 recoge las categorías que establecen y las propuestas curriculares.

3.2.2. Perspectiva evolutiva del talento matemático

La evaluación se apoya en un diseño longitudinal planeado para seguir a lo largo de medio siglo el desarrollo académico, profesional y personal de varias cohortes. El procedimiento es el siguiente. En el momento de la identificación, 12-13 años de edad, los alumnos cumplimentan los dos tests y un cuestionario. Posteriormente se envía un cuestionario a los padres. A los 17-18 años reciben por correo un nuevo cuestionario y otro más amplio cuando tienen 23-24 años. En algunos casos aplican además otras pruebas.

El marco conceptual que guía la evaluación parte de un modelo de adaptación vocacional, la teoría de ajuste al trabajo, en la que el aprendizaje óptimo y el contexto del trabajo son dos niveles que requieren correspondencia mutua y que denominan como satisfactoriedad y satisfacción. La satisfactoriedad denota el grado de correspondencia entre las capacidades y los requisitos exigidos por un entorno determinado, a saber, el currículo o la ocupación. La satisfacción denota el grado de correspondencia entre las preferencias y los tipos de refuerzos de una ocupación o una trayectoria educativa. La Tabla 1 muestra la riqueza del diseño y recoge datos sobre seis cohortes de alumnos.

TABLA 1
*DISEÑO LONGITUDINAL DEL **SMPY**. COHORTES EN ESTUDIO (LUBINSKI, BENBOW Y SANDERS, 1993)*

Cohorte	N	Fecha identificación	Edad identificación	Criterio en el SAT	Nivel de capacidad
1	2188	972-1974	12-13	Matem. ≥ de 390 Verbal ≥ de 370	1%
2	778	1976-1979	12	1/3 superior de los preseleccionados	0.5%
3	423	1980-1983	<13	Matem. ≥ 700 Verbal (630	0.01%
Grupo de comparación		1983	12	SAT-M+SAT-V ≤ 540	5%
4	≅ 750	1987	12	Matem. ≥ 500 Verbal (430	0.5%

La cohorte número 5 incluye 750 estudiantes matriculados en prestigiosas facultades y escuelas de Estados Unidos en varias carreras científicas; fueron estudiados en 1992 cuando tenían 23-25 años.

Cada cohorte está separada unos cinco años de la siguiente y en conjunto ocupan un espacio de veinte años. Están siendo observadas a la misma edad para controlar efectos como el desarrollo y la maduración. En el análisis de datos contemplan varios

grupos formados por nivel de capacidad a los 12-13 años y género o educación acelerada y no acelerada, entre otros.

Sus principales hallazgos son (Stanley y Benbow, 1983; Benbow y Minor, 1986; Swiatek y Benbow, 1991; Benbow, 1991, 1992): a) Mostrar la validez predictiva del SAT de nivel superior en el diagnostico del talento matemático; b) Probar que la alta capacidad permite a los que la poseen manejar con desenvoltura los desafíos intelectuales y académicos que plantea la aceleración del currículo; c) Los chicos obtienen puntuaciones significativamente más altas que las chicas en el SAT-M, desaparecen las diferencias en el SAT-V, se mantiene debilitada la tendencia a una mayor participación masculina en programas de este tipo y estudios más recientes rompen parcialmente estereotipos ligados al género, en el sentido de que las chicas que deciden participar en cursos de aceleración, obtienen en ellos un rendimiento similar al de los chicos, y desaparecen las diferencias significativas tradicionalmente existentes entre los géneros en actitud hacia programas de este tipo. Ambos grupos obtienen altos rendimientos universitarios; el masculino obtiene grados más altos. El seguimiento de la primera cohorte a lo largo de diez años pone de manifiesto que menos del 1% de las chicas situadas en la cola del 1% superior en capacidad, ha proseguido estudios de doctorado en ciencias o ingeniería mientras lo están haciendo ocho veces más chicos (Cfr. Jiménez Fernández, 2000, cap. IV)[3].

4. EL ENFOQUE TRANSVERSAL EN LA EVALUACIÓN DE PROGRAMAS

Este enfoque mide a los sujetos en un momento temporal determinado y estas observaciones frecuentemente se toman como resultados de determinadas intervenciones o programas. Sólo puede medir directamente diferencias interindividuales. Buena parte de los programas que informan de resultados de la evaluación caen dentro de este enfoque, si bien en muchos casos las medidas son tan débiles, que es difícil

---

3 En el caso del género es esclarecedor seguir la evolución e interpretación de las diferencias; la polémica acaecida entre Stanley (Stanley y Benbow, 1983) y Alenxander y Pallas (1983) a raíz de la publicación de un artículo de Benbow y Stanley que revisaba los diez primeros años del SMPY; la contraréplica (Benbow y Stanley, 1983), la finura de datos e interpretaciones más profundas (Lubinski, Benbow y Sander, 1993). Stanley y Benbow mantuvieron en la contrarreplica lo esencial de sus conclusiones y precisaron que se referían a estudiantes del 3% o del 1% superior de la población; que en el nivel de capacidad media la ejecución podría obedecer a otro patrón. Hay autores que se preguntan cómo explicar la convergencia en las puntuaciones medias de los grupos generales mientras permanecen pronunciadas en el caso de los superdotados. Avanzan dos explicaciones, el *método de construcción* de los tests y el *grado de dispersión* de las medias en cada género. El meta-análisis de Friedman (1989) sobre un centenar de investigaciones con muestras generales de alumnos de enseñanza primaria y secundaria concluye que la *diferencia media* entre los géneros en capacidad matemática es muy pequeña y ha decrecido desde 1974 a 1987, intervalo abarcado por el meta-análisis. Análisis de los modelos de socialización, expectativas, opciones curriculares o miedo al éxito muestran que las chicas no están fracasando en los estudios científicos. Simplemente eligen más otras opciones porque les atraen más, exigen menos dedicación y son más fáciles de gerenciar con responsabilidades femeninas futuras; muestran asimismo intereses más amplios que los chicos altamente dotados y lo materializan, por ejemplo, eligiendo cursos de matemáticas y de lengua en la misma proporción mientras los chicos eligen seis veces más cursos del primer tipo que del segundo. Por otra parte, las que deciden entrar en el campo del número, tendencia en aumento, obtienen en él altos rendimientos.

atribuir los efectos al programa. Esbozamos la evaluación de un programa señero, el de Renzulli, del que no parece exista evaluación longitudinal propiamente dicha.

4.1. Evaluación del programa del triple enriquecimiento de Renzulli

El autor del SEM (*Schoolwide Enrichment Model*) concibe la superdotación como el resultado de la convergencia de tres factores: capacidad intelectual superior a la media, motivación de logro y creatividad. El diagnóstico se materializa dando un peso similar a los tests y a las estimaciones del profesorado y la intervención educativa que desarrolla es el enriquecimiento del currículo en aula especial, con tres niveles de complejidad. En el primero se trata de abrir el currículo a contenidos más amplios y estimulantes; el segundo pone el énfasis en la adquicisión de destrezas y estrategias de alto nivel que propicien la elaboración de productos complejos; en el tercero se trata de que el alumno muestre la eficacia del programa realizando productos propios o casi propios de un profesional experto (Renzulli y Reis, 1991, 1994). A diferencia del SMPY es un modelo bien conocido en España.

Las evaluaciones provienen de escuelas, niveles, diseños y aplicaciones de distinta duración y grado de control. Un resumen de los resultados puede ser que (Renzulli y Reis, 1991, 1994; MacSheffrey y Hoge, 1992): a) Su aplicación ejerce un efecto favorable en las prácticas de enseñanza del profesorado y contribuye a mejorar su actitud hacia los más capaces; b) Mejora la motivación y creatividad de los alumnos y los estimula a realizar productos más sofisticados y diversificados; c) Se asocia a cambios positivos en los centros escolares en aspectos relacionados con las prácticas de enseñanza y los proyectos de los alumnos; d) Es interesante para los alumnos de minorías étnicas y con dificultades de aprendizaje, según algún estudio.

5. EL AGRUPAMIENTO POR CAPACIDAD

Bajo este epígrafe se resumen resultados de estudios longitudinales y transversales. Este modelo puede aplicarse con carácter fijo o a tiempo parcial; en el aula ordinaria o en aula separada; para un materia o para todo el currículo y agrupar a los estudiantes de un mismo curso o de distintos cursos. Son realmente variantes del enriquecimiento curricular, goza de tradición en la individualización de la enseñanza y de algún modo tiene mala prensa sin que la investigación más rigurosa respalde tal estereotipo.

De estudios de síntesis se puede concluir (Kulick y Kulick, 1991; Slavin, 1990; Rogers y Span, 1993) que: 1) El agrupamiento por capacidad a tiempo completo no produce ganancias en el rendimiento medio de los alumnos medios y de capacidad baja; sí produce ganancias sustanciales en el caso de los más capaces. 2) Los alumnos de secundaria agrupados en programas de alta capacidad tienen aspiraciones académicas más altas y es más probable que terminen matriculándose en la universidad. Sin embargo, es difícil determinar que sea un efecto claro del agrupamiento. La investigación tampoco ha sido capaz de probar que los profesores de los más capaces sean mejores profesores aunque sí se dan en ellos una serie de circunstancias que les llevan a creer en su trabajo y a proyectarlo en sus clases. 3) El agrupamiento par-

cial para enriquecimiento del currículo produce en los más capaces ganancias generales en el rendimiento, pensamiento crítico y creatividad. 4) El agrupamiento por capacidad a tiempo completo como práctica habitual o con fines de enriquecimiento dentro del aula, tiene escaso efecto en el autoconcepto, aunque al inicio de dicho agrupamiento se produce un ligero descenso de la autoestima de los más capaces que vuelve a reequilibrarse. 5) El agrupamiento por capacidad de los más capaces produce una moderada mejora de la actitud hacia las materias en los que son agrupados. 6) Las distintas formas de agrupamiento del modelo son beneficiosas para los más capaces, no tanto por el agrupamiento en sí como por lo que transcurre en el seno del grupo[4].

6. A MODO DE VALORACIÓN

Señalar en primer lugar, que la visión presentada pretende sensibilizar sobre un tema olvidado en nuestro país y que las condiciones sociales y legislativas han puesto sobre el tapete. Los datos aportados dan una visión objetiva propia de la investigación y deben tener peso en las decisiones de política educativa y práctica escolar, pero no deben ser los únicos. Tienen limitaciones. Por otra parte, educar a la persona es pensar en la distancia corta y en el largo plazo; en el yo, en el nosotros y en su corresponsabilidad. Sin embargo, olvidarlos a su suerte y no formalizar su educación en el marco de la escuela pública dentro de la atención a la diversidad, es ir contra la igualdad de oportunidades, apostar por una mediocridad y en el extremo un fracaso que no merecen y renunciar a una educación para la excelencia de cada alumno. La escuela sin adjetivos debe diferenciar parcialmente su educación.

BIBLIOGRAFÍA

Alexander, K.L. y Pallas, A.M. (1983). Reply to Benbow and Stanley. *American Educational Research Journal*, v.20 (4), 475-478.

Benbow, C.P. (1991). Mathematically talented Children: Can Acceleration Meet Their Educational Needs?, en Colangelo, N. y Davis, G.A. (Ed.), o.c., 154-165.

— (1992). Academic Achievement in mathematics and Science of students Between ages 13 and 23: Are There Differences Among Students in the Top One percent of Mathematical Ability? *Journal of ducational psychology*, v. 84 (1), 51-61.

4 Puede ser de interés resumir que respecto del trabajo cooperativo o agrupamiento heterogéneo se puede concluir que: 1) No se desprende que este tipo de aprendizaje sea académicamente beneficioso para los más capaces. 2) Aunque hay evidencia para apoyar en la población general efectos académicos positivos del aprendizaje cooperativo que incorpora competitividad o responsabilidades individuales formales; no puede extrapolarse esta conclusión a los más capaces, si bien aún no ha sido estudiado en profundidad. 3) Aunque existe alguna evidencia para apoyar resultados afectivos «medibles» derivados del agrupamiento heterogéneo, en particular para la aceptación de la diversidad cultural y de los alumnos objeto de integración escolar, no existen estudios que permitan extrapolar esta conclusión a la población de los superdotados (Rogers y Span, 1993).

— y Stanley, J.C. (1983). Differential Course-taking Hypothesis Revisited. *American Educational Research Journal*, v. 20 (4), 469-573.
— y Minor, L.L. (1986). Mathematically Talented Males and Females and Achevement in the High School Sciences. *American Educational Research Journal*, v. 23 (3), 425-436.
Callahan, C.M. (1993). Evaluation Program and pprocedures for Gifted Education: International Problem and Solutions, en Heller, K.A.; Mönks, F.J. y Passow, A.H. (Ed.), o.c., 605-618.
— (1998). Lesson learned from avaluating programs for the gifted promising practices. *EDUCACIÓN XXI*, 1, 53-72.
— y Caldwell, M.S. (1995). *Practitioner's Guide to Evaluating Programs for the Gifted*. Washington: National Association for Gifted Children.
Cohn, S.J. (1991). Talent Searches, en Colangelo, N. y Davis, G.A. (Ed.), o.c.,167-177.
Colangelo, D. y Davis, G.A. (Ed.) (1991). *Handbook of Gifted Education*. Massachustts: Allyn and Bacon.
Davis, G.A. y Rimm, S.B. (1998). *Education of the Gifted and Talented*. Boston: Allyn and Bacon.
Friedman, L. (1989). Mathematics and the gender: a meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research*, v. 59 (2), 185-213.
Heller, K.A.; Mönks, F.J y Passow, A.H. (Ed.) (1993). *International Handbook of Research and Development of Gifteness and Talent*. Oxford: Pergamon.
Jiménez Fernández, C. (1997). Educación de los alumnos más dotados. *Revista de Investigación Educativa*, v. 15 (2), 217-234.
— (2000), *Diagnóstico y educación de los más capaces*. UNED-MEC. Madrid. Colección, Varia. 404 pp.
Kulick, J. A. y Kulick, CH. (1991). Ability Grouping and Gifted Student, en Colangelo y Davis, (Eds.), o.c., 178-192.
Laylock, B. (1979). *Gifted Children*, Illinois: Scott Foresman.
Lubinsky, D., Benbow, C. y Sander, CH. (1993). Reconceptualizing Gender differences in Achievement Among the Gifted, en Heller, Mónks y Passow (Eds.), o.c., 693-707.
Oden, M.H. (1968). The fullfillement of promise: 40 year follow-up of the Terman gifted group. *Genetic Psychological Monographs*, 77, 3-93.
Renzulli, J.S. y Reis, S.M. (1991). The Schoolwide Enrichment Model: A Comprehensive Plan for the development of creative Productivity», en Colangelo, N. y Davis, G.A. (Eds.). o.c., 111-141.
— (1994), Research Related to the Schoolwilde Enrichment Triad Model. *Gifted Chil Quarterly*, v. 38 (1), 7-20.
Rogers, K.B. y Span, P. (1993). Ability Grouping With Gifted and Talented Students: Research and Guidelines, en Heller, K.A.; Mönks, F.J. y Passow, A.H. (Ed.) (1993). o.c., 585-592.
Slavin, R.E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, 471-499.
Stanley, J.C. (1973). Accelereting the educational progress of intellectually gifted youths. *Educational Psychologist*, 10, 133-146.

548

— y Benbow, C.P. (1983). SMPY´S first decade: ten-years of posing probllems and solving them, *The Journal of Special Education*, v. 17 (1), 11-25.

Subontnik, R.F. y Arnold, K.D. (1993). Longitudinal Studies of Giftedness: Investigating the Fulfillment of Promise, en Heller, K.A.; Mönks, F.J y Passow, A.H. (Ed.) (1993), o.c., 149-160.

Swiatek, M.A. y Benbow, C.P. (1991). Ten-Year Longitudinal Follow-Up of Ability-Matched Acceleratd and Unaccelerated Gifted Students. *Journal of Educational Psychology*, v. 83 (4), 528-538.

Terman, L.M. (1925). *Genetic studies of genius*. California: Stanford University Press.

— y Oden, M.H. (1947). The gifted child grows up. Twenty five years follow up of a superior group. *Genetic studies of genius*, v. 4.

Vantassel-Baska, J. (1992). *Planing Effective Curriculum Gifted Learners*. Denver, Colorado: Love Publishing Company.

Vialle, W. (1994). «Termanal» Science? The Work of Lewis Terman Revisited. *Roeper Review*, v. 17 (1), 32-38.

547

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 565-585

EVALUACIÓN DE PROGRAMAS PARA ALUMNOS DE ALTA CAPACIDAD: ALGUNOS PROBLEMAS METODOLÓGICOS

Javier Tourón

Departamento de Educación
Universidad de Navarra

«Evaluation is the tool of defensibility.
Where this tool is used skillfully, programs
for the gifted will survive» (SEELEY, 1986)

RESUMEN

Este trabajo parte de la premisa de que es necesaria una educación diferencia para los alumnos de alta capacidad, y por tanto el desarrollo de programas que atiendan a las necesidades peculiares de estas personas. La investigación, principalmente llevada a cabo en países de habla inglesa, ha venido mostrando evidencias abundantísimas de esta necesidad.

Los programas se enfrentan con una problemática evaluativa peculiar, pues deben constantemente mostrar que son eficaces para poder subsistir. La evaluación es la garantía para la defensa de estos programas allá donde existen.

Algunos de los problemas metodológicos con los que la evaluación se enfrenta han sido objeto de consideración. De modo particular los relacionados con la medida y el diseño. Problemas, por otra parte, que si bien tienen una peculiaridad propia en los programas dirigidos a una población particular, comparten muchos aspectos con la problemática metodológica general. La metodología de la evaluación, aunque presenta problemas que hemos tratado también es cierto que ofrece pautas razonablemente adecuadas para valorar la eficacia de tales programas.

ABSTRACT

This papar is based on the premise that highly gifted students require a differentiated education and thus the need for the development of programs geared toward the special needs of these

students. Research carried out principally in English-speaking countries offers abundant evidence in this regard.

Programmes are faced with a peculiar set of evaluative problems since, if they are to be continued, their effectiveness must be constantly demonstrated. Ongoing evaluation is a guarantee of the continued support of such programmes, wherever they might be in use.

Some of the methodological problems encountered in evaluation are considered, particularly those related to measurement and design. On the other hand, although problems of a unique kind are frequently found in those programmes designed for a specific population, they invariably have many characteristics in common with the general methodological issues. Although the methodology of evaluation presents problems which we have analysed, it is also true that it offers reasonably adequate guidelines for evaluating the effectiveness of such programmes.

1. INTRODUCCIÓN

Antes de entrar en la consideración de los principales problemas metodológicos que presenta la evaluación de programas dirigidos a personas de alta capacidad, quisiera plantear algunas cuestiones previas, comenzando por la necesidad misma de dichos programas. ¿Son realmente necesarios los programas específicos para estos alumnos? ¿Por qué?

La LOGSE plantea en sus artículos 36 y 37 la importancia de atender a las necesidades educativas permanentes o transitorias derivadas de las peculiaridades de los aprendices. El decreto 696/1995 plantea, por primera vez en la historia reciente de la legislación educativa, que existen necesidades educativas especiales asociadas a sobredotación intelectual y regula las condiciones y procedimientos para flexibilizar (aunque tímidamente) el sistema educativo. Por otra parte, uno de los pilares de la LOGSE es la atención a la diversidad, que lamentablemente se ha entendido con demasiada frecuencia como atención a los déficit de los escolares, pero nunca a los superávits. Cuando se analiza todo el entramado legislativo actual de nuestro país para la enseñanza no universitaria y los principios y características psicopedagógicas que lo animan, es fácil ver que se adapta, o pretende hacerlo, a las necesidades de todos los escolares. Sobre este particular hemos escrito extensamente en otro lugar (Cfr. Tourón y cols. 1998).

Sin embargo, no son pocos los mitos que planean sobre los alumnos superdotados (de alta capacidad), que actúan como freno para su adecuada atención en la escuela. Mitos y estereotipos que están bien lejos de la realidad y necesidades de estos alumnos (Cfr. Tourón y Reyero, 2000). La superdotación es un constructo complejo, multidimensional que ha de verse como una capacidad potencial que necesita de unas determinadas condiciones para que pueda desarrollarse de modo satisfactorio (Gagné, 1993). Están lejos los tiempos en los que superdotación se veía como algo fijo, dado e inmutable, para dar paso a un claro cambio de paradigma en el que la superdotación sólo llegará a desarrollarse si se dan las condiciones adecuadas (Reyero y Tourón, 2000). Como señalan con acierto Treffinger y Feldhusen (1996) «los talentos emergen y crecen evolutivamente, y para algunos no llegan a emerger porque no se produce una adecuada estimulación en la escuela y la familia. Es imperativo que todos los que tra-

549

bajan con jóvenes vean los talentos y potencialidades como algo educable y emergente, y no como algo fijo e inmutable».

La identificación de los alumnos de alta capacidad, paso previo para poder plantear estrategias educativas de intervención (programas), se hace precisa en el momento en el que los programas educativos regulares de la escuela no pueden responder a las demandas educativas de estos alumnos, poniendo en riesgo su desarrollo tanto cognitivo como afectivo. Siguiendo a Feldhusen (1986), hay tres premisas básicas que deben ser tenidas en cuenta para entender la necesidad de una educación diferenciada para los alumnos más capaces, en realidad para todos los alumnos: a) cada estudiante tiene derecho a una educación que sea adecuada a sus especiales características y necesidades; b) todo estudiante tiene derecho a unos servicios educativos que le ayuden a desarrollar sus habilidades potenciales al más alto nivel y c) debemos desarrollar los talentos de los jóvenes de modo que sirvan a las necesidades de su propia nación.

En suma que, a mi juicio, está fuera de duda que la escuela y los programas regulares que en ella se desarrollan, orientados al alumno medio, no pueden dar respuesta adecuada a las demandas psicoeducativas de los alumnos más capaces, por lo que es preciso tender a un sistema educativo más adaptativo que favorezca el desarrollo del potencial de cada escolar. Existen excelentes trabajos en los que se analizan con profusión las características y principios que deben seguirse en el desarrollo de programas diferenciados para los alumnos de alta capacidad, y que aunque éste no es el lugar para tratar (Cfr. Brennam, 1988; Kaplan, 1979; Maker, 1982, 1995; Renzulli, 1995; Van Tassel-Baska, 1984, por citar sólo unos pocos), podemos brevemente apuntar que deben seguir, de acuerdo con el *Leadership Training Institute* (ver Maker, 1986) siete grandes principios:

a) El contenido debe estar enfocado y organizado de modo que permita un estudio más elaborado, complejo y profundo de las principales ideas, problemas y temas que integran el conocimiento en los diversos sistemas de pensamiento.
b) Debe permitir el desarrollo y la aplicación de destrezas de pensamiento productivo que permitan a los estudiantes reconceptualizar el conocimiento existente o producir otro nuevo.
c) Debe permitir explorar los cambios constantes del conocimiento y la información y desarrollar la actitud de que es valioso seguir tales cambios en un mundo abierto.
d) Debe estimular el uso, la selección y exposición de recursos especializados.
e) Debe promover la iniciativa personal y el aprendizaje autodirijido.
f) Debe fomentar la comprensión de uno mismo y de nuestras relaciones con las personas, la sociedad, las instituciones, la naturaleza y la cultura.
g) La evaluación de los programas para alumnos de alta capacidad debe, de acuerdo con los principios anteriores, centrarse en destrezas de pensamiento de alto nivel, creatividad y excelencia en el rendimiento y los productos.

Así pues, el análisis de la investigación tanto teórica como experimental, nos permite señalar, sin muchas dudas, que los movimientos en contra de la superdotación que se aprecian en algunos países, el nuestro no es una excepción, son más producto

550

de posiciones ideológicas concretas que de argumentos educativos y resultados de investigación sólidos que los avalen.

2. LA EVALUACIÓN DE PROGRAMAS PARA ALUMNOS DE ALTA CAPACIDAD

Como señala Seeley (1986) «mientras es cierto que los programas para los superdotados tienen algunas características únicas, no hay necesidad de crear enfoques completamente nuevos para evaluarlos (...). Hay excelentes prácticas de evaluación educativa que son perfectamente adecuadas para los programas de superdotados. No hace falta reinventar la rueda. Lo que no significa que las características especiales de estos programas deban ignorarse» (p. 265). La evaluación de los programas para alumnos de alta capacidad puede llevarse a cabo desde los enfoques denominados tradicionales, vinculados con concepciones *positivistas* de la realidad, de la ciencia y de la evaluación, criticados por muchos como inservibles para captar la verdadera esencia de la acción educativa, o desde concepciones que se agrupan bajo la rúbrica *postpositivista*, son los modelos naturalistas (Cfr. Borland, 1990; Lincoln y Guba, 1985), que conciben la realidad como algo construido, múltiple, donde lo observado se ve en interacción con el observador; donde la generalización se ve como algo imposible y quizá indeseable; y donde el proceso mismo no pretende ser objetivo ni libre la influencia de los valores. Los diseños fijos dan paso a los emergentes, los instrumentos de medida rigurosamente validados dejan su lugar al investigador como principal instrumento de recogida de datos. Lo nomotético es sustituido por lo idiográfico, donde los datos no son *descubiertos*, como si estuviesen ahí fuera, sino que son literalmente *creados* (Guba y Lincoln, 1989). Son modelos que al oponerse a los de corte positivo se denominan alternativos (Cfr. Dinham y Udall, 1986; Callahan y Cadwell, 1986). Pero no es el propósito de este trabajo entrar en el análisis de los modelos posibles para evaluar programas para alumnos de alta capacidad, sino analizar algunos de los problemas metodológicos que en dicha evaluación se plantean. La discusión sobre los enfoques evaluativos ya ha sido objeto de tratamiento en otro lugar de este número monográfico.

Sí que es preciso decir que el análisis de los problemas metodológicos que más adelante vamos a realizar se alinea con una óptica cuantitativa de la evaluación, compatible por otra parte con enfoques y modelos muy diversos.

La evaluación de programas se ha visto vinculada al movimiento de rendición de cuentas nacido en los EEUU a mediados del siglo pasado (para una visión comprensiva general puede consultarse Tejedor, 1994; García Ramos, 1992) y si bien esta es una razonable función, no siempre es la que mayor impacto positivo tiene en la mejora de los mismos. Sin embargo, por los problemas metodológicos que analizaremos enseguida, y por otras razones, la evaluación ha sido con frecuencia vista como una amenaza de supresión de programas especiales, que se ven exigidos a mostrar unos resultados y una eficacia que rara vez se pide para los programas regulares. La evaluación se reduce a un carácter sumativo que, si bien es importante, es netamente incompleto. Como señala Borland (1997, p. 255), «la mejora de los programas es uno de los resultados más importantes de la evaluación y puede ser una de las razones más poderosas para llevar a cabo el proceso.

551

(...) Dicha mejora debe ser uno de nuestros imperativos, de modo que forme parte de nuestra concepción general de lo que deben ser los propósitos de la evaluación».

Según Renzulli (1975) la evaluación de programas para los superdotados tiene que cumplir cinco propósitos: a) Descubrir si los objetivos se han cumplido o no y en qué grado; b) descubrir consecuencias inesperadas y no planeadas derivadas de las prácticas del programa; c) determinar las políticas subyacentes y las actividades relacionadas que contribuyen al éxito o fracaso en áreas particulares; d) ofrecer un continuo *feedback* durante el proceso en etapas intermedias a lo largo del programa y d) sugerir cursos de acción alternativos, reales e ideales, para modificar el programa.

Callahan (1993) refiriéndose a la importancia de la evaluación de los programas señala seis aspectos clave que deben tenerse en cuenta si pretendemos tener procedimientos de intervención defendibles: a) la evaluación debe entenderse como una parte integrante del diseño y planificación del programa; b) los problemas que surgen en la evaluación de los programas no pueden ser causa que justifique los fallos de la evaluación; c) la evaluación como proceso está cambiando tanto en sus propósitos como en su amplitud; d) la evaluación no supone sólo determinar el valor de un programa; e) los nuevos desarrollos de la evaluación pueden ser de utilidad en la evaluación de los programas para superdotados y f) la evaluación acaba siendo lo que se quiere que sea.

No es posible, sin embargo llevar a cabo una adecuada evaluación de un programa sin una adecuada descripción del mismo, sin un adecuado establecimiento de los estándares, de los puntos de referencia con los que comparar, sin un plan operativo, sin una adecuada base para atribuir los resultados, sin una previsión de cómo actuar ante los posibles problemas que puedan surgir.

No obstante, la literatura especializada en este campo abunda en una serie de problemas y resistencias con las que se encuentra la evaluación de programas. Las principales dificultades se pueden agrupar en torno a nueve aspectos que hemos reelaborado, siguiendo a Callahan (1993), del siguiente modo:

1) La evaluación se ve como una amenaza
2) Con frecuencia los programas están mal definidos y descritos
3) Existen dificultades para determinar cuál es 'el programa' para poder aislar sus efectos
4) No siempre se formulan las preguntas de evaluación apropiadas ni se establecen adecuadamente las prioridades en la evaluación
5) La comparación de los efectos del programa con determinados estándares y el establecimiento de los grupos de control es difícil
6) El profesor como programa
7) Falta de atención a las posibles interacciones entre aptitud y tratamiento
8) Poca claridad en el establecimiento de los indicadores de éxito y problemas en la instrumentación
9) La utilización de la evaluación (sumativa, formativa, administrativa, etc).

Vamos a analizar ahora algunos de los principales problemas de carácter metodológico que surgen en la evaluación de estos programas, principalmente desde una óptica cuantitativa.

552

3. ALGUNOS PROBLEMAS METODOLÓGICOS EN LA EVALUACIÓN DE PROGRAMAS PARA ALUMNOS DE ALTA CAPACIDAD

Dadas las limitaciones de espacio disponible, vamos a seleccionar sólo algunos de los problemas que consideramos más importantes. Los vamos a organizar de acuerdo a cuatro apartados: a) problemas derivados de la concepción de superdotación; b) las metas y objetivos del programa; c) problemas de medida y d) problemas con los diseños de evaluación. Trataremos los dos primeros con más brevedad y nos extenderemos algo más en los dos últimos por tener una relación más directa con las cuestiones estrictamente metodológicas.

a) *La concepción de superdotación*

Son múltiples las concepciones tanto implícitas como explícitas que se han propuesto sobre la superdotación (Cfr. Sternberg y Davidson, 1986), algunas de las cuales pueden verse desarrolladas en Tourón y cols. (1998); así mismo, en los últimos años se ha venido produciendo una clara modificación del paradigma clásico hacia un nuevo paradigma más centrado en la identificación y desarrollo del talento (un tratamiento extenso puede verse en Reyero y Tourón, 2000). Pero estas diversas concepciones no son un problema *per se* respecto a la evaluación de los programas, ya que como señala Carter (1991), los evaluadores pueden operacionalizar un determinado concepto de superdotación y analizar los resultados del programa en función de lo que se espera a partir del concepto adoptado. Lo que realmente constituye un problema es que los responsables del programa (*stakeholders*) pueden tener diferentes concepciones de la superdotación y esperar resultados diversos del programa, con lo cual es difícil llegar a una evaluación de la bondad del mismo.

«Los evaluadores deben cerciorarse de que los *stakeholders* están trabajando desde el mismo marco de referencia que los evaluadores, tienen las mismas expectativas y están de acuerdo respecto al tipo de estudiantes a los que el programa está sirviendo» (Carter, 1991, p. 249), y por tanto, hay un marco común de referencia y un acuerdo previo sobre los resultados que se esperan como efecto del desarrollo del programa. No es necesario que el evaluador y los responsables del programa coincidan en el concepto de superdotación, lo que es preciso es que se pongan de acuerdo en el concepto que va a operar en un programa dado y de qué modo se operacionalizará. A partir de ahí será más fácil ponerse de acuerdo en las metas y objetivos en los que debe centrarse la evaluación.

b) *Las metas y objetivos*

Este es otro problema potencial de la evaluación. No es infrecuente que las metas y objetivos estén formulados de una manera vaga o ambigua. Y como señala Borland (1997, p. 257) «las metas y objetivos que no especifican claramente qué se espera que los alumnos ganen como resultado de la existencia del programa son de escasa utilidad en la evaluación». Algunos autores recomiendan que la evaluación se centre en

grandes metas como: el incremento de la creatividad, la capacidad de resolución de problemas, el fomento de estrategias de pensamiento, etc. Pero hay otro tipo de resultados más concretos y medibles que suelen venir reflejados en los objetivos de tipo curricular que el programa persigue. Algunos autores recomiendan que la evaluación se centre en las primeras y abogan por el desarrollo de diseños de evaluación que permitan hacerlo (Cfr. Gallagher, 1979), no obstante es fácil ver la complicación que entraña el determinar cuáles serán los indicadores que se considerarán válidos para metas tan genéricas. Por otro lado se puede producir un efecto negativo evidente y es que al existir una relativa distancia entre los indicadores y las metas, la información que se obtenga de la evaluación puede ser poco útil para mejorar el programa, o bien que no se pueda llevar a cabo una atribución razonable entre el programa y el efecto producido.

Por ello, y sin perder de vista la importancia de las grandes metas, al servicio de las cuales deben estar los objetivos más específicos (metas intermedias), la evaluación debe centrarse en buena parte sobre éstos. Los resultados específicos podrán ser utilizados con carácter formativo y podrán emplearse en la mejora y modificación del programa. La solución óptima es llegar a un adecuado compromiso entre la evaluación de las grandes metas y los objetivos curriculares específicos que se supone tienden a ellas.

c) *Problemas de medida*

La evaluación de programas, particularmente desde un enfoque cuantitativo entraña serios problemas relacionados con la medición, no siempre relacionada con el uso tests, aunque éstos son los que presentan los problemas más complejos, tanto si hablamos de la medición de los resultados como de variables de entrada, de contexto o de proceso. Vamos a apuntar algunos de estos problemas.

Quizá el primero de ellos sea la falta de instrumentos adecuados de la que muchos autores se hacen eco (Cfr. Borland, 1997). Este problema bastante generalizado en muchos contextos se manifiesta de modo particularmente grave en el nuestro, donde es difícil encontrar procesos sistemáticos de desarrollo y validación de instrumentos, así como de actualización de los existentes (Cfr. Tourón, Repáraz y Peralta, 1999).

Siguiendo a Feldhusen y Jarwan (1993), podemos señalar entre los criterios clásicos para la adecuada elección de los instrumentos: a) relevancia del test, b) fiabilidad, c) validez, d) baremación, e) sesgos posibles y f) efecto de techo (para un tratamiento de algunos de los señalados puede consultarse Martínez Arias, 1995 y Muñiz, 1996. Y naturalmente es obligada la lectura de los *Standards for Educational and Psychological Testing*, 1999).

Aunque casi todos ellos son muy obvios merecen un comentario en esta panorámica general, ya que si bien pueden considerarse problemas generales de cualquier tipo de evaluación, los problemas son mayores cuando hablamos de una población tan específica como la de los alumnos de alta capacidad, para la que la ausencia de instrumentos adecuados es casi general.

La relevancia del test se refiere a la adecuación entre el propósito para el cual ha sido diseñado y el uso que se pretende hacer de él. Por ejemplo, si pretendemos seleccionar

los candidatos más adecuados para un programa de desarrollo de la capacidad matemática, un test de inteligencia general no parece lo más adecuado; del mismo modo un test de habilidad matemática tiene poco sentido —aunque sea técnicamente correcto— si se pretende seleccionar sujetos para un programa de desarrollo de la creatividad en artes plásticas. Por tanto, al hablar de relevancia estamos refiriéndonos a la adecuación del test para el propósito específico para el que se va a utilizar. Naturalmente este problema está relacionado con la decisión que se tome de evaluar resultados generales (metas) o específicos (objetivos) en un determinado programa. Cierto es que se hace preciso no perder de vista otros procedimientos para abordar la estimación de los efectos de los programas sin el uso de test. Sería el caso de utilizar otras modalidades de evaluación (*assessment*) como el *consensual assessment* utilizado para el estudio de la creatividad, por citar sólo una alternativa (Cfr. Amabile, 1983).

La fiabilidad no precisa demasiados comentarios. Se trata de una condición esencial, aunque no suficiente, para que un test pueda ser empleado en un proceso de evaluación. Es importante valorar la información disponible sobre la fiabilidad de la prueba que pensemos utilizar: sobre qué muestras se ha obtenido, con qué procedimientos, hace cuánto tiempo, etc. Asimismo, relacionado con la fiabilidad, será importante hacer uso del error de medida, ya que permitirá realizar juicios más precisos sobre las puntuaciones individuales, el establecimiento de intervalos de confianza, puntos de corte, etc. Estos datos son importantes a la hora de tomar decisiones.

Ahora bien, todo lo señalado se refiere a una perspectiva de la medida desde la óptica de la teoría clásica, cuyas limitaciones son suficientes como para que se consideren otros abordajes más acordes con los desarrollos modernos de la misma. Nos referimos a la TRI (Teoría de Respuesta al Item) que permite superar muchas de las limitaciones de la teoría clásica. Lamentablemente no es posible extenderse en este punto ahora, pero baste señalar que la TRI debería tenerse más en cuenta en los procesos de evaluación (Cfr. Orden y cols., 1998; Tourón y Gaviria, 2000a y b), toda vez que nos permite aplicar modelos de tests adaptativos computerizados (o no) por ejemplo, de modo que se maximiza la información que se puede obtener de un sujeto con un 'gasto' mínimo de recursos, ya que los ítems que se le presentan se adaptan a su competencia. Se evita así que un sujeto se vea obligado a responder ítems demasiado fáciles o difíciles para él, lo que en el caso que nos ocupa es crucial. Este tipo de estrategias maximizan la información que se puede obtener y ofrecen un error específico para cada puntuación estimada, lo cual es bastante más plausible que calcular un error común para todas las puntuaciones como se hace en la teoría clásica.

Así pues, y dado que el tratamiento de este tema está fuera de las posibilidades de este breve trabajo, se puede señalar que, además de las aportaciones de tipo técnico que ofrecerá la TRI a la hora de construir tests y aplicarlos a situaciones concretas de evaluación, por ejemplo, «su gran contribución se centra en la posibilidad de obtener mediciones invariantes respecto de los instrumentos utilizados y de los sujetos implicados. En la TCT el resultado de la medición de una variable depende del test utilizado (...). En la Teoría Clásica la medición de una variable es inseparable del instrumento utilizado para medirla y ello constituye una seria limitación, pues inevitablemente se acabará definiendo operativamente la variable por el instrumento con que se mide (...).

Además, las propiedades del instrumento de medida, esto es, de los ítems y, por tanto, del test, están en función de los sujetos a los que se aplican (...). El acercamiento clásico se encontraba encerrado en esa incongruencia teórica: la medición depende del instrumento utilizado y las propiedades de éstos están en función de los objetos medidos, de los sujetos. El objetivo central de la TRI será solucionar este problema» (Muñiz, 1990).

La validez es la *condicio sine qua non*. Un modo clásico sencillo de referirse a la validez es decir que se trata de una apreciación del grado en el que un instrumento mide aquello que pretende. Más precisamente habría que decir que la validez no es tanto del instrumento aunque está implicado, —naturalmente— cuanto de las inferencias que pretendamos hacer a partir de las puntuaciones del mismo. Es conocido que clásicamente hemos distinguido entre diversos tipos de validez: de contenido, concurrente, predictiva, convergente, discriminante, etc., pero la concepción más inclusiva de todas ellas es la validez de constructo, que supone una inserción de la medida en la teoría, de modo que medir se convierte en una forma de validar una teoría, la estructura teórica del fenómeno medido. Pero como recomiendan los *Standards* de 1999, citados más arriba, es más correcto hablar de diversos tipos o fuentes de evidencia sobre la validez que de diferentes tipos de validez. «La validez es un concepto unitario. Es el grado en el que la evidencia acumulada apoya las interpretaciones pretendidas para el uso del test. Como los *Standards* de 1985 esta edición se refiere a tipos de evidencias respecto a la validez, más que a diferentes tipos de validez» (*Standards*, 1999).

Sin entrar en mayores tecnicismos ahora, podemos señalar que se trata aquí, para los propósitos que perseguimos, de responder a dos preguntas: a) ¿qué constructo queremos medir?, b) ¿qué evidencias muestra este instrumento de ser una medida adecuada de este constructo? No parece necesario insistir en la importancia de esta característica, sin la cual todas las demás son superfluas (Cfr. APA, 1986; Cronbach, 1970; Cronbach y Meehl, 1955; Tourón, 1989).

Los baremos son una pieza de información imprescindible para poder interpretar las puntuaciones de un determinado test. Para determinar el grado de excepcionalidad y rareza (Cfr. Sternberg, 1993; Sternberg y Zhang, 1995) de las competencias o talentos de una determinada persona es preciso compararla con sujetos comunes en alguna característica, generalmente la edad, el nivel escolar, etc. Pues bien, los baremos de un test lo que reflejan es el comportamiento típico de un grupo concreto en el test, es decir, su nivel de ejecución. Así pues, no será posible decir cuan excelente o rara es una determinada capacidad sin conocer qué es lo esperable en sujetos de esa edad, por ejemplo.

Su importancia es capital, ya que sin baremos adecuados no podremos, desde una perspectiva normativa, valorar el grado o nivel de ejecución de un sujeto en la prueba correspondiente y por tanto será difícil, sino imposible estimar los efectos del programa que queramos evaluar. Un ejemplo patente de este problema, y sus implicaciones en el proceso de identificación (extensible a la evaluación de programas) puede verse en Tourón, Repáraz y Peralta, (1999). Por eso abordar procesos de validación rigurosos que aporten baremos actualizados obtenidos sobre muestras actuales y suficientemente representativas es esencial (pueden consultarse a este respecto los trabajos que venimos realizando de baremación del SCAT en Navarra, por ejemplo, Tourón y cols., 2000; Tourón, 2000).

556

Los efectos de sesgo son otro de los criterios a tener en cuenta a la hora de seleccionar un instrumento de medida. Los sesgos se refieren, entre otras cosas, al hecho de que las puntuaciones obtenidas por los sujetos pueden ser inferiores o, en general, verse alteradas, por razón de su sexo, raza, situación cultural, religión, etc., lo que llevaría a una inadecuada valoración de los mismos. El sesgo, como señalan Feldhusen y Jarwan (1993), es —principalmente— un problema de fiabilidad del diagnóstico. La justicia (adecuación) del diagnóstico es una cuestión de validez. Por ejemplo, sería poco razonable someter a los alumnos españoles a un test de razonamiento verbal en el que muchos items incluyesen vocabulario perteneciente a algún deporte típicamente norteamericano, como el beisbol o el fútbol americano. Del mismo modo, sujetos que hayan vivido en el ámbito rural extremo durante toda su vida tendrán problemas para contestar a tests profundamente impregnados de cultura urbana. Todos estos efectos producen sesgos que llevan a los sujetos a obtener puntuaciones que no reflejan su habilidad o capacidad real en la variable medida. Por lo mismo, un test de inteligencia general excesivamente verbalizado producirá un sesgo claro en sujetos deficientemente escolarizados o que viven en un ámbito culturalmente deprivado.

Los problemas de sesgo se analizan modernamente a partir del estudio del funcionamiento diferencial de los ítems (*diferential item functioning*). Pero conviene no confundir ambos aspectos. En efecto, un ítem puede tener un funcionamiento diferencial, para digamos chicos y chicas y no tener sesgo. La definición de DIF aclarará este extremo.

«Se dice que un ítem funciona diferencialmente para dos o más grupos si la probabilidad de dar una respuesta correcta a un determinado ítem está asociada con la pertenencia de sujetos de la misma capacidad a uno de los grupos. Si el grado de DIF es significativo desde el punto de vista práctico y puede ser atribuido plausiblemente a una característica del ítem que es relevante para el constructo medido, entonces la presencia de este ítem en el test sesga la estimación de la habilidad de algunos individuos» (Holland y Wainer, 1993).

Así pues, para planificar el proceso de evaluación será necesario atender a la validez y equidad del test para la población específica para la que se va a emplear, al tiempo que se deben estudiar con cautela los baremos disponibles y todas las evidencias que el constructor del test pueda ofrecer respecto al uso e interpretación de las puntuaciones del mismo.

El efecto de techo es el último de los aspectos que queremos señalar en relación con la medida, pero en absoluto el menos importante. Más aún, es un aspecto crítico. «Se refiere, como es sabido, a la falta de un rango de dificultad adecuado en los ítems, lo que conduce a que los sujetos más capaces no puedan demostrar adecuadamente todo su potencial. Dicho en otros términos, el test pierde la capacidad de discriminar o distinguir las diferencias entre los sujetos a partir de determinado nivel. De este modo, cuando se produce el efecto de techo, sujetos muy distintos en su potencial aparecerán como iguales al obtener puntuaciones similares» (Tourón y cols. 1998).

Utilizar un test que no presenta una dificultad adecuada para los sujetos más competentes en un ámbito dado es como hacer una carrera de 100 metros lisos para descubrir corredores de fondo. Ciertamente todos llegarán a la meta, pero si detenemos ahí la

carrera, nunca sabremos a donde podrían haber llegado los corredores con mejor forma física, cuáles realmente son corredores de fondo. Y lo que es peor, consideraremos a todos como velocistas, cuando muchos de ellos no lo son. Esto es particularmente serio a la hora de valorar los efectos de un programa, pues si el test o los tests que empleemos no tienen suficiente *recorrido*, aparecerán como iguales alumnos de competencias muy diversas. Por otra parte, para complicar más las cosas, lo veremos al hablar del diseño, si estos instrumentos se utilizan como medidas pretest y posttest, los alumnos más aventajados tendrán pocas o ninguna posibilidad de mostrar sus ganancias como resultado del programa, pues ya tenderán a obtener puntuaciones muy altas en el pretest. Además, unido a esto está el conocido efecto de regresión por el que los sujetos que en un pretest toman posiciones muy altas tenderán a obtener puntuaciones más bajas en segundas medidas con el mismo test. Este efecto puede atenuar o llegar a cancelar efectos del programa que son reales. Este es un *artefacto* estadístico que debe tenerse muy presente y que afecta a la validez interna de los diseños (Campbell y Stanley, 1979) y que puede paliarse utilizando diferentes instrumentos para las medidas pre o postest, lo que vuelve poner en primer plano la problemática de la medida.

Este efecto será tanto más grave, lógicamente, cuanto más extremos sean los sujetos evaluados. Se considera que comienza a presentarse este efecto cuando la puntuación media de un grupo está por encima del 75% de la puntuación máxima del test, o cuando la distribución de las puntuaciones está muy sesgada negativamente.

Uno de los mejores sistemas para corregir el efecto de techo es utilizar el procedimiento denominado en el ámbito sajón «*out of level testing*», es decir, utilizar tests previstos para sujetos de mayor edad que la de aquéllos que van a ser evaluados (Cfr. Feldhusen, 1991). Este es un sistema utilizado con probado éxito en el estudio de la precocidad matemática a partir del modelo denominado Talent Search (Cfr. Benbow, 1991; Stanley, 1991; Tourón y Reyero, en prensa).

d) *Problemas con los diseños de evaluación*[1]

Así como los problemas tratados anteriormente pueden ser comunes a evaluaciones centradas en el contexto, el proceso, las variables de entrada, etc., los problemas de diseño que vamos a tratar brevemente se refieren principalmente a evaluaciones del producto, de resultados. Cualquier evaluación de programas está relacionada de un modo u otro con la evaluación de los resultados obtenidos por aquéllos alumnos que ha recibido dicho programa. Ordinariamente los resultados han de compararse con los de otro grupo de sujetos de las mismas características pero que no han estado sometidos al efecto del programa. Es, como se comprende, la estrategia clásica del diseño experimental en la que es ocioso entrar aquí. Lo que sí puede tener interés, por ser una problemática importante en la aplicación de este modelo, es el llamado problema del *grupo de comparación* y el problema del *control*. El primero relacionado,

1 En el anexo I incluimos un cuadro con los diseños pre-experimentales, quasiexperimentales y experimentales más frecuentes en la evaluación de programas, analizando las limitaciones a su validez interna y externa de acuerdo con la obra de Campbell y Stanley.

como es obvio, con la selección adecuada de un grupo de comparación para el que recibe el programa (tratamiento), el segundo relacionado con la compleja problemática del control de las variables dentro del diseño. Estos problemas y otros, que no son del caso, han llevado a muchos a preferir el abandono del modelo experimental por costoso e inadecuado a la realidad educativa y a optar por diseños (o paradigmas) alternativos. A nuestro juicio, la evidencia experimental y su aproximación a la causalidad no pueden ser aparcadas por razones más ideológicas que científicas. La capacidad probatoria del diseño está muy por encima de supuestas metodologías más *flexibles*, aunque su puesta en práctica pueda representar serios problemas, algunos de los cuales vamos a analizar.

Aunque sea sucintamente señalemos que los grupos de comparación son grupos de «control no equivalentes» (intactos) porque no se han formado por procedimientos aleatorios, pero que se consideran suficientemente equiparables a los grupos experimentales, y por tanto no son grupos de control en sentido estricto, según la terminología clásica del diseño (Winner, 1971). Cuando los procedimientos de formación de los grupos de comparación no son aleatorios se nos plantean una serie de problemas que vamos a analizar. Desde luego el mejor grupo de comparación para un grupo de niños de alta capacidad que reciben un programa de resolución de problemas, por ejemplo, sería aquél formado por niños de alta capacidad de su mismo entorno que no han recibido el programa. Esto plantea problemas bien obvios, tanto políticos como éticos, y pocos padres estarían dispuestos a que sus niños fuesen privados de una ayuda potencialmente beneficiosas para ellos. Veremos alternativas a este problema.

Desde la lógica del diseño y atendiendo a la validez interna del mismo (Campbell & Stanley, 1966), es difícil poder atribuir los efectos de un programa (tratamiento) sin una comparación estricta con un grupo de control formado aleatoriamente. Aunque irónicamente, según señala Carter (1991) citando a Snow (1974), un diseño que opere sobre grupos aleatorizados puede dejar de ser un *diseño representativo* del contexto del programa que pretendamos evaluar. Por eso Snow recomienda diseños que representen el contexto natural donde los escolares se desenvuelven normalmente y no los diseños artificiales que pueden llevar a los sujetos actuar de modo diferente a cómo lo harían de no estar sujetos a la *manipulación experimental*. El problema es ser capaces de establecer un *equilibrio aceptable* entre las exigencias de la evaluación del efecto producido por el programa y la naturalidad del contexto. Este es el dilema clásico entre la investigación de campo y la de laboratorio. Como señala Carter (1991, p. 262): «Esto es por lo que los evaluadores deben seleccionar diseños que estén lo más próximos posible al diseño ideal, mientras que se acomodan a las restricciones y circunstancias de la situación. Sea cual fuere el diseño que se elija debe permitir al evaluador responder a las preguntas clave de la evaluación en el tiempo asignado al proyecto».

En relación con los grupos de control no equivalentes, que hemos *llamado grupo de comparación*, se han propuesto algunas soluciones que no están exentas de problemas, pero que vamos a comentar brevemente. La primera de ellas es la *equiparación*. Se trataría de seleccionar escuelas o distritos escolares que pudiesen ser emparejados en determinadas variables consideradas relevantes para la igualación de los grupos. Este procedimiento que es viable en determinados contextos tiene problemas evidentes, no

obstante, ya que es difícil determinar en qué variables se deben equiparar los grupos, y más difícil todavía determinar que cualquier otra variable no considerada no pueda convertirse en una hipótesis rival alternativa al efecto del programa. Por otra parte es improbable que un distrito o escuela haya identificado a alumnos de alta capacidad y no haya establecido algún tipo de programa para ellos. Más aún, aun aceptando que es posible equiparar al grupo experimental con un grupo de control razonablemente igualado a él en algunas variables, muchas otras quedarán necesariamente fuera de control, con lo que cualquier atribución causal del efecto del programa estará comprometida.

Un diseño bastante interesante propuesto por Callahan (1983) para la evaluación de programas está inspirado en la técnica del contrabalanceo y está particularmente diseñado para resolver el problema del grupo de control. En el cuadro 1 reproducimos un esquema del mismo con una modificación importante introducida por Carter (1991). En este diseño se comparan alumnos de alta capacidad que han recibido el programa con otros que no, pero para ello se divide el currículo en unidades que se aplican de modo alterno a los grupos. Así, mientras en el momento 1 un grupo recibe la

CUADRO 1
DISEÑO PROPUESTO POR CALLAHAN Y MODIFICADO POR CARTER PARA LA EVALUACIÓN DEL CURRÍCULO PARA ALUMNOS DE ALTA CAPACIDAD

Grupos	Evaluación de una Unidad curricular	
	Tiempo 1	Tiempo 2
Grupo A (alta capacidad)		
G1	X	Y
G2	X	Y
G3	X	Y
Grupo B (alta capacidad, comparación)		
G4	Y	X
G5	Y	X
G6	Y	X
Grupo C (no alta capacidad)		
R1	X	Y
R2	X	Y
R3	X	Y
Grupo B (no alta capacidad, comparación)		
R4	Y	X
R5	Y	X
R6	Y	X

X: unidad curricular experimental; Y= otra unidad curricular

unidad X, el otro recibe la unidad Y, siendo la situación contraria en el momento 2. De este modo el grupo expuesto a la unidad X actúa de control para el grupo que ha recibido la unidad Y, y del mismo modo, el Y será control para el que ha recibido la unidad X. Obsérvese que se trata de alumnos de alta capacidad en ambos casos que han sido divididos (aleatoriamente de modo ideal) en dos grupos, con lo que la equivalencia está garantizada. Hay algunos problemas prácticos en los que no vamos a entrar, pero que están relacionados con las posibilidades de compartimentar el currículo de este modo o de alterar las secuencias de las unidades, por citar sólo dos casos evidentes.

Pero veamos la segunda parte del cuadro, que es análoga a la primera pero en este caso las unidades se alternan con grupos de sujetos no de alta capacidad. Los grupos C y D reciben las mismas unidades desarrolladas por los profesores de los grupos A y B. De este modo el grupo A es comparado con el C (ambos reciben la misma unidad X) y el grupo B es comparado con el D (ambos reciben la unidad Y). De esta forma, según Carter (1991) se puede estudiar el efecto de la educación diferenciada para los superdotados en variables como la tasa de aprendizaje, la amplitud y la profundidad, ya que se estarán comparando clases regulares con clases de alta capacidad. Esto permitirá a los evaluadores comprobar si las clases regulares pueden beneficiarse del currículo desarrollado para los superdotados en el mismo grado que éstos.

En ocasiones es imposible disponer de un grupo de comparación, por limitaciones de la propia realidad educativa en la que vamos a evaluar un programa o porque no es posible establecer un grupo comparable. Aún en este caso es importante poder establecer algún tipo de comparación que permita valorar el programa.

Un procedimiento que se ha descrito en la literatura es el *pretest retrospectivo*. En este caso los sujetos actúan como su propio control. Una vez que los sujetos han recibido la instrucción su rendimiento es evaluado con algún test o cuestionario que se considere apropiado. A continuación se les ofrece el mismo instrumento y se les pide que respondan al mismo como lo habrían hecho antes de recibir la instrucción. Los resultados se comparan para analizar las posibles diferencias. La debilidad del procedimiento reside en la confianza que se pueda conceder a la capacidad de los sujetos para autoevaluarse en este modo. Payne y Browne (1982) citado por Carter (1991) ofrecen resultados satisfactorios sobre todo en el campo de variables afectivas, aunque afirman haberlo utilizado con éxito también en el campo cognitivo.

Analicemos para terminar este, necesariamente rápido, repaso por algunos de los problemas más importantes en la evaluación de programas, la problemática del control cuando las limitaciones del contexto impiden la manipulación de variables, la formación aleatoria de grupos, etc. Hay tres grupos de diseños comúnmente utilizados: el causal comparativo, los diseños correlacionales y los diseños cuasiexperimentales. Veamos algunas particularidades de ellos.

Los *diseños causales comparativos* son diseños que se utilizan cuando la manipulación de variables es imposible. En su forma más simple se trata de dos grupos naturales (intactos), uno de los cuales ha recibido el programa y el otro no. La ausencia de control es total, pues ni siquiera es posible determinar aleatoriamente qué grupo recibirá el programa. El único control que puede ejercer el evaluador es la selección del

grupo de comparación, que deberá ser tan similar al que ha recibido el programa como sea posible. Aunque los análisis estadísticos que se llevan a cabo con estos diseños son iguales que los que se realizan con diseños experimentales, es preciso no cometer el error de hacer el mismo tipo de inferencias. Sin control sobre las variables toda inferencia causal será inadecuada, sólo será posible realizar afirmaciones relacionales. La debilidad de este diseño es patente, aunque ofrece un grupo de comparación que de ser probada su equivalencia en variables relevantes para el programa, aporta una cierta información.

Los *diseños correlacionales* son una alternativa interesante aunque poco utilizada por los evaluadores. Difieren de los anteriores en que sólo utilizan un grupo y los datos se analizan con técnicas correlacionales. Naturalmente estos diseños utilizan procedimientos que van más allá de la correlación bivariada, empleándose con frecuencia procedimientos de regresión múltiple, análisis discriminante o correlación canónica. Cuando se utiliza la regresión múltiple, la variable dependiente es el resultado que se pretende medir en el programa (simple o múltiple) y las variables dependientes son aquéllas cuya capacidad predictiva sobre tal resultado interesa analizar. De este modo es posible obtener información que, aunque no puede ser causal, es muy interesante para valorar distintas dimensiones del programa. Ni que decir tiene que las posibilidades de los modelos causales o los recientes desarrollos del análisis multinivel permitirán avances muy sustantivos en la evaluación de programas.

Para terminar esta sucinta exposición, podemos hacer mención a algunos diseños cuasi experimentales, quizá los que mejor se adaptan a las limitaciones y condicionamientos contextuales a los que debe atender la evaluación. Si bien es cierto que distan de los verdaderos experimentos en cuanto al control, no cabe duda de que ofrecen una información muy valiosa en la evaluación al tiempo que controlan muchas de las amenazas a la validez interna (ver anexo I).

Dos de estos diseños son: el diseño de *series cronológicas de grupo único* y el diseño de *series cronológicas grupo de control no equivalente*[2].

El primero de ellos utiliza un sólo grupo al que se mide a intervalos regulares antes del tratamiento (programa) y después de terminar la intervención. El disponer de medidas antes y después permitirá determinar el posible efecto del programa, para ello será preciso analizar la serie temporal de medidas antes y después y comprobar la tendencia y el cambio producido por el programa, lo que no puede hacerse de modo correcto en un diseño en el que sólo tuviésemos una medida antes y otra después. Las limitaciones de espacio no nos permiten analizar ejemplos ilustrativos de este diseño (puede verse la obra de Campbell y Stanley, 1966, o Fitz-Gibbon y Morris, 1987).

Finalmente, el diseño de series temporales con grupo de control no equivalente en su forma más simple consiste en un desarrollo del anterior pero utilizando dos grupos naturales (pueden verse las amenazas a la validez en el anexo I). El trata-

---

2 Un diseño muy interesante utilizado con éxito en la evaluación de programas es el de discontinuidad en la regresión, que aquí no vamos a tratar. Puede verse una descripción en Stanley y Robinson (1986) y una aplicación práctica en Robinson y Stanley (1989).

miento es aleatoriamente asignado a uno de los grupos actuando el otro como control (más bien como comparación, pues al no haber sido asignados los sujetos aleatoriamente, no es adecuado denominarlo de control). Como señala Carter (1991, p. 268), «la no equivalencia de los grupos es la mayor debilidad de este diseño. Para interpretar correctamente los efectos del tratamiento, los evaluadores deben detectar las diferencias entre los grupos antes del tratamiento a partir del los pretests y de comparaciones en otras variables que puedan ser relevantes para el programa. Si los análisis en las variables pretest no revelan diferencias significativas, los resultados del programa se puede analizar a partir de la comparación entre los posttests, pero si las hubiese, la utilización de un control estadístico, como el análisis de varianza sería necesario, en realidad es común y conveniente aplicar este control aunque no se encuentren diferencias significativas, siempre que las variables (covariantes) medidas antes del programa sean realmente relevantes y tengan un posible impacto en los resultados».

4. CONCLUSIONES

En las páginas precedentes hemos tratado de plantear la justificación a una educación diferenciada para los alumnos de alta capacidad. La investigación, principalmente llevada a cabo en países de habla inglesa, ha venido mostrando evidencias abundantísimas de esta necesidad. Es pueril pensar que los niños de alta capacidad se desarrollarán adecuadamente sin una intervención y ayuda adecuadas. Simplemente no es así, pero no es ésta una cuestión de opinión, sino de resultados de investigación convenientemente contrastados.

A pesar de ello, determinados prejuicios llevan a muchas autoridades educativas y a los educadores mismos a presentar resistencias más o menos fuertes en contra de una educación diferenciada, que no segregada (el principio de integración tampoco lo permitiría). Por esta razón, los programas se enfrentan con una problemática evaluativa peculiar, pues deben constantemente mostrar que son eficaces para poder subsistir. La evaluación, como señalamos al comienzo, es la garantía para la defensa de estos programas allá donde existen.

Algunos de los problemas metodológicos con los que la evaluación se enfrenta han sido objeto de consideración. De modo particular los relacionados con la medida y el diseño. Problemas, por otra parte, que si bien tienen una peculiaridad propia en los programas dirigidos a una población particular, comparten muchos aspectos con la problemática metodológica general.

La situación de nuestro país en lo que al desarrollo de programas para alumnos de alta capacidad se refiere es muy deficiente, por lo que los resultados de evaluación son prácticamente inexistentes, al menos hasta donde conocemos. O si existen no llegan a ser publicados en los canales ordinarios de divulgación.

Es preciso, a nuestro juicio, flexibilizar el sistema educativo y conseguir una escuela más adaptativa que ofrezca a cada alumno las ayudas que precisa para su desarrollo. Al mismo tiempo se hace patente la necesidad de diseñar programas específicos para atender a tales necesidades, más allá de lo que la escuela regular ofrece. La meto-

dología de la evaluación, aunque presenta problemas que hemos tratado también es cierto que ofrece pautas razonablemente adecuadas para valorar la eficacia de tales programas, al menos desde la óptica que aquí se ha adoptado.

BIBLIOGRAFÍA

Amabile, T.M. (1983). *The Social Psychology of Creativity*. New York: Springer-Verlag.

American Educational Research Association (1985). *Standards for Educational and Psychological Testing*. Washington: AERA, APA, NCME.

American Educational Research Association (1999). *Standards for Educational and Psychological Testing*. Washington: AERA, APA, NCME.

Benbow, C.P. (1991). Mathematical Talented Children: Can Acceleration meet their Educational Needs? En N. Colangelo. y G.A. Davis (Eds.). *Handbook of Gifted Education*.

Borland, J.H. (1990). Postpositivist Inquiry: Implications of the «New Philosophy of Science» for the Field of the Education of the Gifted. *Gifted Child Quarterly*, 34, 161-167.

Borland, J.H. (1997). Evaluating Gifted Programs. En N. Colangelo y G. A. Davis (Eds.). *Handbook of Gifted Education*. 2nd Ed. Boston: Allyn & Bacon.

Brennam, W. (1988). *El curriculo para niños con necesidades especiales*. Madrid: M.E.C. Siglo XXI.

Callahan, C.M. (1983). Issues in Evaluation Programs form the Gifted. *Gifted Child Quarterly*, 27, 33-37.

Callahan, C.M. (1993). Evaluation Programs and Procedures for Gifted Education: International Problems and Solutions. En K.A. Heller, F.J. Mönks y A.H. Passow (1993). *International Handbook of Research and Development of Giftedness and Talent* (pp. 605-618). Oxford: Pergamon Press.

Callahan, C.M. y Caldwell, M.S. (1986). Defensible Evaluation of Programs for the Gifted and Talented. En J. Maker. *Critical Issues in Gifted Education. Defensible Programs for the Gifted*. Volume I (pp. 277-296). Austin: Pro-Ed.

Campbell, D.T. y Stanley, J.C. (1966). Experimental and Quasi-experimental Designs for Research. *Chicago*: Rand McNally.

Campbell, D.T. y Stanley, J.C. (1973). *Diseños experimentales y quasi experimentales en la investigación social*. Buenos Aires: Amorrortu.

Carter, K.R. (1991). Evaluation of Gifted Programs. En N.K. Buchanan y J.F. Feldhusen (Eds.). *Conducting Research and Evaluation in Gifted Education. A Handbook of Methods and Applications* (pp. 245-272). New York: Teachers College Press.

Cronbach, L.J. (1970). Test Validation. En Thorndike, R. L. (Ed.). *Educational Measurement*. Washington: American Journal of Education.

Cronbach, L.J. y Meehl, P.E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, 52, 281-302.

Dinham, S.M. y Udall, A.J. (1986). Evaluation for Gifted Education: Synthesis and Discussion. En J. Maker. *Critical Issues in Gifted Education. Defensible Programs for the Gifted*. Volume I (pp. 297-316). Austin: Pro-Ed.

Feldhusen, F.J. (1991). Identification of Gifted and Talented Youth. En Wang, M.C.; Reynolds, M.C. y Walberg, H.J. (Eds.). *Handbook of Special Education*. Vol 4. Oxford: Pergamon Press.

Feldhusen, J.F. (1986).Policies and Procedures for the Development of Defensible Programs for the Gifted. En J. Maker. *Critical Issues in Gifted Education. Defensible Programs for the Gifted*. Volume I (pp. 235-256). Austin: Pro-Ed.

Feldhusen, J.F. y Jarwan, F.A. (1993). Identification of Gifted and Talented Youth for Educational Programs. En K.A. Heller, F.J. Mönks y A.H. Passow (1993). *International Handbook of Research and Development of Giftedness and Talent* (pp. 233-251). Oxford: Pergamon Press.

Fitz-Gibbon, C.T. y Morris, L.L. (1987). How to design a program evaluation. Newbury Park: Sage.

Gagné, F. (1993). Constructs and Models Pertaining to Exceptional Human Abilities. En K.A. Heller, F.J. Mönks y A.H. Passow (1993). *International Handbook of Research and Development of Giftedness and Talent*. Oxford: Pergamon Press.

García Ramos, J.M. (1992). Recursos metodológicos en la evaluación de programas. *Bordón*, 43(4), 461-476.

Gallagher, J.J. (1979). Research Needs for the Education of the Gifted. En J.J. Gallagher, J.C. Gowan, A.H. Passow y E.P. Torrance (Eds.). *Issues in Gifted Education* (pp. 79-91). Ventura, CA: Ventura County Superintendent of Schools.

Guba, E.G. y Lincoln, Y.S. (1989). Fourth Generation Evaluation. Beverly Hills, CA: Sage.

Kaplan, S.N. (1979). *Insevice Training Manual: Activities for Developing Curriculum for the Gifted/Talented*. Los Angeles: Leadership Training Institute on the Gifted and Talented.

Lincoln, Y.S. y Guba, E.G. (1985). Naturalistic Inquairy. Beverly Hills, CA: Sage.

Maker, C.J., & Nielson, A.B. (1995). *Curriculum Development and Teaching Strategies for Gifted Learners*. Boston: Allyn and Bacon.

Maker, C.J. (1982). *Curriculum Development for the Gifted. Rockville*, MD: Aspen Systems.

Maker, C.J. (1986). Defensible Programs for Gifted Students: What are they? En J. Maker. *Critical Issues in Gifted Education. Defensible Programs for the Gifted*. Volume I (pp. 277-296). Austin: Pro-Ed.

Martínez Arias, R. (1995). Psicometría: Teoría de los tests psicológicos y educativos. Madrid: Síntesis.

Muñiz, J. (Coord.)(1996). *Psicometría*. Madrid: Editorial Universitas.

Muñiz, J. (1990). Teoría de Respuesta a los Items. Un nuevo enfoque en la evolución psicológica y educativa. Madrid: Pirámide.

Orden Hoz, A.; Bisquerra, R.; Gaviria, J.L.; Gil, G; Jornet, J. López, F. Sánchez, J.; Sánchez, M.C.; Sierra, J. y Tourón, J. (1998). *Los resultados escolares. Diagnóstico del Sistema Educativo 1997*. Madrid: INCE, Ministerio de Educación y Cultura.

Payne, D.A. y Brown, D.L. (1982). The use and Abuse of Control Groups in Program Evaluation. *Roeper Review*, 5, 11-14.

Renzulli, J.S. (1975). *A Guidebook for Evaluating Programs for the Gifted and Talented*. Ventura, CA: Office of the Ventura County Superintendent of Schools.

565

Renzulli, J.S. (1995). Intervenciones educativas para el desarrollo de la superdotación en los niños. Ponencia presentada en el II Congreso Internacional de Psicología y Educación. Madrid, 16-18 noviembre.

Reyero, M., y Tourón, J. (2000). Reflexiones en torno al concepto de superdotación: evolución de un paradigma. *Revista Española de Pedagogía*, 215, pp. 7-38.

Robinson, A. y Stanley, T.D. (1989). Teaching to Talent: Evaluating and Enrich and Accelerated Mathematics Program. *Journal for the Education of the Gifted*, 12(4), 253-267.

Seeley, K.R. (1986). Evaluation for Defensible Programs for the Gifted. En J. Maker. *Critical Issues in Gifted Education. Defensible Programs for the Gifted*. Volume I (pp. 265-277). Austin: Pro-Ed.

Snow, R.E. (1974). Representative and Quasi-representative Designs for Research on Teaching. *Review of Educational Research*, 44, 265-291.

Stanley, T.D. y Robinson, A. (1986). Regression Discontinuity: Integrating Research and Program Design in Programs for the Gifted. *Journal for the Education of the Gifted*, 9(3), 181-191.

Stanley, J.C. (1996). In the Beginning: The Study of Mathematical Precocious Youth. En C.P. Benbow y D. Lubisnki (Eds.). *Intellectual Talent. Psychometric and Social Issues* (pp. 225-235). Baltimore, MD: The Johns Hopkins University Press.

Sternberg, R.J. (1993). Procedures for Identifying Intellectual Potencial in the Gifted: A Perspective on Alternative «Metaphors of Mind». En Heller, K.A.; Mönks, F.J. y Passow, A.H. (Eds.). *International Handbook of Research and Development of Giftedness and Talent*. Oxford: Pergamon Press.

Sternberg, R.J. y Zhang, L. (1995). What do We Mean by Giftedness? A Pentagonal Implicit Theory. *Gifted Child Quarterly*, 39(2), 88-94.

Sternberg, R.J., & Davidson, J.E. (1986). *Conceptions of Giftedness*. Cambridge: Cambridge University Press.

Tejedor, F.J., García-Valcárcel, A. y Rodríguez Conde, N.J. (1994). Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo. *Revista de Investigación Educativa*, 23, 93-127.

Tourón, J. (2000). Expanding the Talent Search in Spain. The Validation of the School and College Ability Test in Spain: Comparison of two Pilot Studies. Symposium paper presented at the 7th ECHA Conferencie. Debrece (Hungary). Agoust, 18-22.

Tourón, J. y Gaviria J.L. (2000a). *Evaluación de la educación primaria en la Comunidad Foral de Navarra*. Pamplona: Dirección General de Educación. Gobierno Foral.

Tourón, J. y Gaviria J.L. (2000b). *Evaluación de la educación primaria en la Comunidad de la Rioja*. Pamplona: Dirección General de Educación. Gobierno de la Rioja.

Tourón, J. y Reyero, M. (2000). Mitos y realidades en torno a la superdotación. En L. Almeida; E.P. Oliveira y A.S. Melo (Coords.). *Alunos sobredotados: contributos para a sua identificaçao e apoio* (pp. 19-27). Braga, Portugal: ANEIS.

Tourón, J. y Reyero, M. (en prensa). La identificación de alumnos de alta capacidad un reto pendiente para el sistema educativo. Madrid: XII Congreso de Pedagogía.

Tourón, J.; Peralta, F., y Repáraz, C. (1998). *La superdotación intelectual. Modelos, identificación y estrategias educativas*. Pamplona: EUNSA.

566

Tourón, J.; Repáraz, Ch. y Peralta, F. (1999). The Identification of High Ability Students: results of a detection process in Navarra (Spain). *High Ability Studies*, 10(2), 163-181.

Tourón, J.; Repáraz, C.; Peralta, F.; Gavria, J.L.; Fernández, R.; Ramos, J.M. y Reyero, M. (2000). La validación del SCAT (*School and College Ability Test*) en Navarra: resultados del estudio piloto. En L. Almeida; E.P. Oliveira y A.S. Melo (Coords.). *Alunos sobredotados: contributos para a sua identificaçao e apoio* (pp. 81-97). Braga, Portugal: ANEIS.

Trefinger, D.J. y Feldhusen, J.F. (1996). Talent Recognition and Development: Succesor to gifted Education. *Journal for the Education of the Gifted, 19*(2), 181-193.

Van Tassel Baska, J. (1984). Appropriate Curriculum for the Gifted. En J.F. Feldhusen (Ed.). *Towards Excellence in Gifted Education*. Denver: Love.

Winner, B.J. (1971). *Statistical Principles in Experimental Designs*. New York: MacGraw Hill.

567

VALIDEZ INTERNA Y EXTERNA DE ALGUNOS DISEÑOS UTILIZADOS EN LA EVALUACIÓN DE PROGRAMAS

			EXPERIMENTAL		CUASIEXPERIMENTAL			PREEXPERIM.
			Diseño 1 Grupo de control pretest-posttest	Diseño 2 Grupo de control sólo posttest	Diseño 3 Pretest-posttest con grupo de control no equivalente	Diseño 4 Series cronológicas grupo único	Diseño 5 Series cronológicas grupo de control no equivalente	Diseño 6 Pretest-posttest de grupo único
			R O X O R O O	R X O R O	O X O O O	OOO X OOO	OOO X OOO OOO OOO	O X O
FUENTES DE INVALIDACION	INTERNA	Historia	+	+	+	—	+	—
		Maduración	+	+	+	+	+	—
		Administración de tests	+	+	+	+	+	—
		Instrumentación	+	+	+	?	+	—
		Regresión	+	+	?	+	+	?
		Selección	+	+	+	+	+	+
		Mortalidad	+	+	+	+	+	+
		Interacción de selecc. maduración, etc.	+	+	—	+	+	—
	EXTERNA	Interacción de admon. de tests y X	—	+	—	—	—	—
		Interacción de selección y X	?	?	?	?	—	—
		Efectos reactivos Interferencia de X múltiples	?	?	?	?	?	?
			(4)	(6)	(10)	(7)	(14)	(2)

Los números corresponden a la ordenación de Campbell y Stanley

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 587-599

EVALUACIÓN DE PROGRAMAS DE EDUCACIÓN EMOCIONAL

Manuel Álvarez, Rafael Bisquerra, Eva Fita, Francesc Martínez y Núria Pérez

CONCEPTO DE EDUCACIÓN EMOCIONAL

Según el art. 1.a de la LOGSE, el sistema educativo se orientará al pleno desarrollo de la personalidad del alumnado. Esto implica atender no solamente al desarrollo cognitivo, sobre el cual hay un énfasis claramente predominante, sino también al desarrollo emocional que lo debe complementar. De las dificultades emocionales pueden derivarse estados de ansiedad, estrés, depresión, consumo de drogas, violencia, conducción temeraria, trastornos de la alimentación, fobia escolar, indisciplina, etc. Diversos trabajos recientes han puesto de manifiesto como el *analfabetismo emocional* tiene unos efectos altamente negativos sobre las personas y sobre la sociedad (Goleman, 1995). De hecho la mayoría de bajas laborales son debidas a causas relacionadas con las emociones (ansiedad, estrés, depresión); por esta misma causa los medicamentos más consumidos son tranquilizantes, ansiolíticos y antidepresivos, con un elevado coste en la seguridad social. La respuesta educativa a esta necesidad social puede ser la educación emocional.

Entendemos por **educación emocional** a un *proceso educativo, continuo y permanente, que pretende desarrollar el conocimiento sobre las propias emociones y las de los demás con objeto de capacitar al individuo para que adopte comportamientos que tengan presente los principios de prevención y desarrollo humano* (Bisquerra, 2000).

Los *objetivos generales* de la educación emocional pueden resumirse en los siguientes términos: a) adquirir un mejor conocimiento de las propias emociones como aspecto importante del autoconcepto; b) identificar las emociones de los demás de cara a facilitar unas mejores relaciones interpersonales; c) desarrollar la habilidad para controlar y canalizar de forma apropiada las propias emociones; d) prevenir los efectos perjudiciales de las emociones negativas intensas (ira, miedo, estrés, depresión, etc.); e) desarrollar la habilidad para experimentar emociones positivas (alegría, humor,

amor, felicidad) y generarlas de forma consciente y voluntaria; f) desarrollar la inteligencia emocional y las competencias emocionales consecuentes; g) desarrollar la habilidad de automotivarse; h) adoptar una actitud positiva ante la vida.

Los *contenidos* de la educación emocional, como consecuencia de lo anterior, giran en torno a temas como los siguientes: Concepto de emoción y tipos de emociones; emoción y salud; emoción y motivación; la inteligencia emocional; conocer las propias emociones; reconocer las emociones de los demás; control emocional; las habilidades socio-emocionales; las emociones en la resolución de conflictos; las emociones en la toma de decisiones; emociones y consumo de drogas; adoptar una actitud positiva ante la vida; bienestar subjetivo y calidad de vida.

Consideramos que la educación emocional es un aspecto más de la orientación psicopedagógica, centrada en la prevención y el desarrollo humano. En otros trabajos (Bisquerra y Álvarez, 1998) se exponen los modelos de intervención en orientación y las estrategias de implantación que de menos a más implicación se concretan en: orientación y educación ocasional, programas en paralelo, asignaturas optativas, acción tutorial, integración curricular y sistemas de programas integrados. Estas mismas estrategias son apropiadas en una intervención de educación emocional.

La educación emocional puede entenderse como un tema transversal en el que participan la totalidad del profesorado en sus clases a lo largo de todo el currículum académico. Las lecciones pueden ser muy breves. Los diversos temas del programa se introducen en un determinado curso pero son retomados, ampliados y profundizados en sucesivos cursos posteriores. No se trata de que cada curso se repita todo, sino de que cada tópico sea contemplado varias veces a lo largo de todo el currículum

LA EVALUACIÓN DE LAS EMOCIONES

Toda intervención en educación emocional, como cualquier acción orientadora, precisa de la evaluación para optimizarse y constatar sus resultados. Sin embargo somos conscientes de la problemática que ello entraña. Las emociones son subjetivas y por tanto difíciles de medir. Como en toda medición, hay que procurar asegurar la validez del instrumento y un elevado coeficiente de fiabilidad. Esto no es fácil al intentar medir las emociones. Pero hay otros aspectos conflictivos señalados por la literatura.

La medición de las emociones debe fundamentarse sobre una teoría de las emociones. De acuerdo con la teoría psicoevolucionista de Plutchik (1991) para medir las emociones hay que medir cada uno de sus componentes (fisiológico, comportamental y cognitivo) por separado. Además sería conveniente medir otros factores asociados como los rasgos de personalidad, estilos de afrontamiento, defensas del yo, etc.

Las mediciones del componente **fisiológico** de las emociones toman como indicadores diversas características como la tasa cardíaca, temperatura de la piel, el ritmo respiratorio, etc. A un nivel más sofisticado se utilizan registros electromiográficos (por ejemplo faciales). Otras variables neurofisiológicas y neuroquímicas se utilizan en psicoendocrinología y psicofarmacología (por ejemplo: niveles de serotonina, noradrenalina y dopamina en relacion con trastornos del estado de ánimo). Las nuevas tecnologías están aportando instrumentos como la TEP (Tomografía por Emisión de

Positrones), la IRMF (Imagen por Resonancia Magnética Funcional) y la MEG (Magnetoencefalografía). Sin embargo, todo este instrumental se aparta bastante de lo que se utiliza habitualmente en educación emocional.

Entre las técnicas utilizadas actualmente en la medición del componente **comportamental** de las emociones están la filmación en vídeo de expresiones para su análisis posterior en cámara lenta[1]. La clave para intuir las emociones consiste en sintonizar con los canales de comunicación no verbal (tono de voz, gestos, expresión facial, etc.). Para evaluar esta habilidad Rosenthal (1977) diseñó el test PONS (Profile Of Nonverbal Sensitivity). La utilización de estos instrumentos, cuyo uso puede ser recomendable en investigación, en la práctica habitual de los programas de intervención resulta como mínimo complicada. No obstante, consideramos que la medición de los comportamientos es importante para la evaluación de programas de educación emocional ya que nos aproxima a la evaluación de la competencia emocional adquirida por el sujeto; en consecuencia, uno de los retos de futuro recae en la elaboración de instrumentos o recursos con tal propósito.

La medida del componente **cognitivo** se basa en la utilización de instrumentos como cuestionarios, escalas o autoinformes estructurados o informales, los cuales son el único recurso disponible para la medición del aspecto cognitivo-subjetivo de las emociones. Algunos de estos instrumentos han logrado una considerable validez y fiabilidad en los últimos años. Sin embargo hay que utilizar sus resultados con cautela por diversas razones que pueden influir en los resultados (memoria, aprendizaje, dominio del lenguaje, pensamientos y creencias sobre las emociones, deseabilidad social, etc.).

Consideramos que en la evaluación de programas de educación emocional a lo que se puede aspirar, en estos momentos, es a la medición del componente cognitivo de las emociones. Esto supone utilizar instrumentos de medida, basados en dos tipologías de instrumentos: a) intrumentos estandarizados, b) instrumentos de carácter más cualitativo. Véase las figuras 2 y 3.

MODELOS DE EVALUACIÓN DE PROGRAMAS APROPIADOS PARA LA EDUCACIÓN EMOCIONAL

Se ha cifrado en más de cincuenta los modelos de evaluación de programas propuestos en los últimos años. Sin embargo, en el ámbito de la orientación, los modelos que han causado mayor impacto son aquellos que se orientan a la toma de decisiones. De entre ellos destacamos en el cuadro siguiente sus rasgos fundamentales.

Para la evaluación de los programas de educación emocional, partiendo de experiencias previas en evaluación de programas, el GROP (Grup de Recerca en Orientació Psicopedagògica) ha optado por tomar como referente el modelo CIPP de Stufflebeam,

---

1 En concreto hay dos métodos de uso predominante: a) FACS (Facial Action Coding System) elaborado por Ekman y Friesen (1978), que se basa en un minucioso análisis molecular de cada uno de los movimientos musculares de la cara considerados de forma independiente; b) MAX (Maximally discriminative facial movement coding system), desarrollado por Izard (1979), con una modificación posterior denominada AFFEX (Izard et al., 1983), que se basa en un análisis molar que implica hacer ciertas inferencias sobre las emociones subyacentes.

571

MODELO/AUTOR	PROCESO/FASES	OBSERVACIONES
Modelo Global de Atkinson, Furlog y Jaroff (1979)	► Recoge información cualitativa sobre el proceso de orientación. ► Recoge información cualitativa sobre los resultados. ► Recoge información cuantitativa sobre las acciones del orientador. ► Recoge información cuantitativa sobre los resultados.	Las diferentes categorías de datos recogidos son útiles a diferentes audiencias. Favorece la autoevaluación del orientador.
Modelo de Alkin (1969)	Se estructura en cinco áreas de decisión: ► Elección del problema (análisis de necesidades i priorización) ► Elección del programa ► Implementación del programa (Evaluación del proceso) ► Mejora del programa (recoge información útil para su optimización) ► Certificación. (Recoge y valora informaciones sobre efectos)	Algunos autores lo han considerado una variante del modelo CIPP
Modelo CIPP de Stufflebeam (1987)	Estructura la evaluación del programa entorno a cuatro contenidos evaluativos: ► Evaluación de contexto (recoge información para la toma de decisiones sobre la implantación y el diseño) ► Evaluación de entrada (pretende tener elementos de juicio que justifiquen el plan de intervención) ► Evaluación de proceso (verificación de la puesta en práctica) ► Evaluación de producto (logros de la intervención que pueden referirse a la eficacia, eficiencia, efectividad y/o impacto del programa)	Los datos recogidos en cada uno de los cuatro contenidos de la evaluación posibilitan la toma de decisiones posterior, según los casos, planificación, programación, implementación y reciclaje.
Modelo de Pérez Juste (1990)	Considera que la evaluación de programas educativos debe realizarse en tres momentos: ► El progra a en sí mismo. (Consiste en el análisis de la calidad intrínseca del programa, de su adecuación a los destinatarios y a la situación de partida) ► El programa en su desarrollo. (Se atiende tanto a la ejecución en sí como al marco contextual de aplicación) ► El programa en sus resultados. (Se atiende no solo a los niveles de logro, incorporando la valoración de los mismos y la continuidad del programa)	Parte del principio de que cualquier actividad puede ser mejorada y es, de hecho mejorable. La evaluación de programas será la base de mejora tanto de los mismos como del personal implicado, y, de modo indirecto, del cuerpo social en que se encuentra inmerso.

Figura 1.
Modelos de evaluación.

al considerar que, a la vez que nos permite reparar en los cuatro contenidos evaluativos antes citados, también nos ofrece la posibilidad de utilizar una amplia diversidad metodológica para la recogida de datos, especialmente cuando se trata de aspectos emocionales de difícil medición.

INSTRUMENTOS PARA LA MEDICIÓN DE LAS EMOCIONES

A continuación se presentan algunos de los instrumentos que pueden ser útiles para medir aspectos cognitivos de las emociones. Algunos de estos instrumentos pueden ser útiles para constatar niveles conseguidos en algunos aspectos concretos de la educación emocional, por ejemplo: el cuestionario de autocontrol infantil y adolescencia (CACIA) puede ser útil en aquellas ocasiones en que sólo se hayan trabajado aspectos relacionados con el control emocional, o el cuestionario de asertividad de M. Dasi y otros (1997) puede ayudar en la evaluación de intervenciones sobre habilidades socioemocionales. En la figura 2 se incluyen instrumentos estandarizados, mientras que en la figura 3 se presentan instrumentos de carácter más cualitativo.

BEST COPY AVAILABLE

DENOMINACIÓN	AUTOR/EDITOR	EDAD	TEMÁTICA	TIPO DE APLICACIÓN	TIEMPO DE APLICACIÓN
CAS Cuestionario de Ansiedad Infantil (B)	J.S. Gillis / TEA	De 6 a 8 años	Conciencia emocional (ansiedad)	Colectiva	Variable, entre 20 y 30 minutos
CDS Escala de Depresión para Niños (B)	M. Lang y M. Tidser / TEA	Entre 8 y 16 años	Conciencia emocional (depresión)	Colectiva	Variable, entre 30 y 40 minutos
CET-DE Cuestionario Estructural Tridimensional para la Depresión (C)	F. Alonso Fernández / TEA	Adolescentes y adultos	Conciencia emocional (estado depresivo)	Individual	Variable, 15 minutos, aprox.
CMAS-R Escala de ansiedad Manifiesta en Niños (Revisada)	Cecil B. Reynold y Bert O. Richmond / TEA	Niños y Adolescentes	Conciencia emocional (ansiedad)	Colectiva	Variable, 15 minutos, aprox.
IDDA-EA Inventario Diferencial de Adjetivos para el Estudio del Estado de Ánimo (B)	J.M. Tous y A. A. Pueyo / TEA	Adolescentes y adultos	Conciencia emocional (activación, estrés y arousal)	Colectiva	Variable, entorno a 10 minutos
ISRA Inventario de Situaciones y Respuestas de Ansiedad	J.J. Miguel Tobal y A. R. Cano Vindel / TEA	A partir de 16 años	Conciencia emocional (ansiedad)	Colectiva	Variable, 30 minutos, aprox.
STAI Cuestionario de Ansiedad Estado Rasgo (B)	C. D. Spielberger, R.L. Gorsuch y R.E. Lushene / TEA	Adolescentes y adultos	Conciencia emocional (estado emocional, situaciones amenazantes, ansiedad)	Colectiva	Variable, 20 minutos, aprox.
STAIC Cuestionario de ansiedad Estado/Rasgo en Niños (B)	C.D. Spielberger / TEA	De 9 a 15 años	Conciencia emocional (ansiedad estado/ansiedad rasgo)	Colectiva	Variable, entre 15 y 20 minutos
BF Cuestionario "Big Five" (B)	G.V. Caprara, C. Barbaranelli y L. Borgogni / TEA	A partir de los 16 años	Control emocional (energía, amigabilidad, consciencia y estabilidad emocional)	Individual y colectiva	Variable, unos 30 minutos
CACIA Cuestionario de Autocontrol Infantil y Adol. (B)	A. Capafons Bonet y F. Silva Moreno / TEA	Niños y adolescentes (11 a 19 años)	Control emocional (autocontrol)	Colectiva	Variable entre 30 y 40 minutos
AFA Autoconcepto Forma A (A)	G. Musitu, F. García y M. Gutiérrez / TEA	De 12 a 18 años	Autoconcepto-Autoestima (académico, social, emocional y familiar)	Colectiva	Variable, en torno a 10 minutos
AF-5 Autoconcepto Forma-5 (A)	G. Musitu y F. García / TEA	Adolescentes y adultos	Autoconcepto-Autoestima (social, académico profesional, emocional, familiar y físico)	Colectiva	Variable, 10 minutos Aprox.
BAS Batería de Socialización 1, 2 y 3 (A)	F. Silva y MªC. Martorell / TEA	De 6 a 15 años	Habilidades socio-emocionales (grado de liderazgo, jovialidad, sensibilidad social y adaptación social)	Colectiva	Variable unos 20 minutos, aprox.
ACS Escalas de Afrontamiento para Adolescentes (A)	E. Frydenberg y R. Lewis / TEA	De 13 a 17 años.	Habilidades socio-emocionales (resolución de problemas)	Colectiva	10 minutos por escala
AECS Actitudes y Estrategias Cognitivas Sociales (A)	M. Moraleda, A. González Galán, A.J. García-Gallo / TEA	Adolescentes de 12 a 17 años (1º y 2º ciclos de la ESO)	Habilidades socio-emocionales (estilo cognitivo, percepción social y estrategias en la resolución de problemas sociales)	Colectiva	Variable, 45 minutos aprox.
EHS Escala de Habilidades Sociales (B)	E. Gismero	Adolescentes y adultos	Habilidades socio-emocionales	Individual o colectiva	Entre 10 y 15 minutos
SIV Cuestionario de Valores Interpersonales (B)	L.V. Gordon / TEA	Adolescentes y adultos	Habilidades socio-emocionales (valores personales, sociales, profesionales y familiares)	Colectiva	Variable, 15 minutos aprox.
EBP Escala de Bienestar Psicológico (B)	J. Sánchez Cánovas / TEA	Adolescentes y Adultos	Habilidades de vida y bienestar subjetivo (bienestar subjetivo, bienestar material, bienestar laboral y bienestar de pareja)	Individual o colectiva	Variable unos 20 Minutos, aprox

Figura 2.
Instrumentos estandarizados.

BEST COPY AVAILABLE

DENOMINACIÓN	AUTOR/EDITOR	EDAD	TEMÁTICA	TIPO DE APLICACIÓN
Bienestar subjetivo (felicidad)	Eysenck (1990)	Adultos	Nivel de felicidad de un sujeto adulto	Individual o colectivo
Comprende su inteligencia emocional	La Vanguardia Magazine (24/11/1996)	Adolescencia Educación secundaria	Inteligencia emocional	Individual o colectivo
¿Le está usted educando de una forma emocionalmente inteligente?	La Vanguardia Magazine (21/06/1998)	Familia	Inteligencia emocional	Individual
¿Cuánto estrés padece?	Adaptación de hindle (1998)	Adultos	Control del estrés	Individual
Test de evaluación asertiva (Adaptación del test de Gambal y Richey)	Güell,M. Y Muñoz,J. (1998)	Adultos	Asertividad	Individual o colectivo
Cuestionario sobre las conductas que utilizas para afrontar las situaciones de estrés	Güell,M. Y Muñoz. J. (1998)	Adolescencia y juventud	Estrés	Individual o colectivo
Autocontrol	Güell,M. Y Muñoz,J	Adolescencia y juventud	Autocontrol	Individual o colectivo
Mide tu desarrollo personal	Revista de Psicología y Salud Natural nº 8	Jóvenes y adultos	Nivel de equilibrio de tu personalidad	Individual
Cuestionario de asertividad	Dasi,M. Y otros (1997)	Adultos	Conducta asertiva	Individual
Tu estado de ánimo	Revista de Psicología y salud natural. Nº6	Adolescentes y adultos	Situaciones de estado de animo	Individual
Si te llevas mal con alguien ¿No será por tu culpa?	Revista Psicología y salud natural nº8	Adolescencia y adultos	Emociones	Individual
Mide tu autodominio	Revista Psicología y salud natural nº6	Adolescencia y adultos	Impulsividad, control de la emotividad y autocontrol	Individual
¿Cuál es tu problema?	Revista Psicología y salud natural nº6	Adolescencia y adultos	Estrés	Individual o colectivo
Inteligencia emocional	Revista Psicología y salud natural nº 13	Adolescentes	Inteligencia emocional	Individual o colectivo
PANAS	Watson,D Clark,L.A. y Tellegen, A. (1998) Versión española Sandin (1995)	Adolescentes	Sentimientos y emociones	Individual o colectivo
Racional o Emotivo	Revista Psicología y salud natural nº 12	Adolescentes	Racional y emocional	Individual o colectivo
EMHAS (Escalas Magallanes de hábitos Asertivos en casa y en el aula)	García,E.M. y Magaz, A. (2000)	12 a 18 años	Hábitos de comportamiento asertivo en casa y en el aula	Individual y colectivo
ADCAs (Escala de actitudes y Valores ante la interacción social)	García,E.M. y Magaz, A. (2000)	Desde los 12 años	Eficacia de un programa de Entrenamiento Asertivo	Individual y colectivo
EVHACOSPI (Test de Evaluación de Habilidades Cognitivas de Solución de Problemas Interpersonales)	García,E.M. y Magaz, A. (1998)	De 4 años en adelante	Habilidades cognitivas de solución de problemas interpersonales	Individual y colectivo
EMANS (Escala Magallanes de Ansiedad)	García,E.M. y Magaz, A. (1998)	Desde los 12 años	Niveles de Ansiedad	Individual y colectivo
EPANS (Escala Magallanes de problemas de Ansiedad)	García,E.M. y Magaz, A. (1998)	De 12 a 18 años	Problemas de ansiedad	Individual y colectivo

Figura 2.
Instrumentos cualitativos.

BEST COPY AVAILABLE

GUÍA PARA EL PROCESO DE EVALUACIÓN

Para llevar a cabo la evaluación de programas de forma operativa, es conveniente seguir una serie de fases para cada uno de los cuatro contenidos evaluados del módulo CIPP (EV. de contexto, EV de entrada, EV de proceso y EV. de producto).

Figura 3.
Fases en la evaluación de programas.

Primera fase: ***Elaboración del plan de evaluación.*** En esta fase se trata de sentar las bases sobre las que se va a fundamentar todo el proceso de evaluación. Para ello, se ha dar respuesta a las siguientes cuestiones:

A) ¿Cuál es el objetivo de la evaluación?

A la hora de afrontar el proceso de evaluación, es tan importante, tener en cuenta los productos como el proceso que ha contribuido a ellos. Así sugerimos:

a) Comprobar el nivel de conocimientos y competencias emocionales adquiridos por los alumnos.
b) Tomar conciencia de los diferentes estados emocionales manifestados por alumnos y profesores.
c) Tener en cuenta las circunstancias favorecedoras o no de la aplicación del programa.

B) ¿Quién recibirá los resultados de la evaluación? Se han de identificar las audiencias que harán uso de los informes la evaluación:

— Alumnado
— Profesorado
— Profesor de Psicología y Pedagogía
— Equipo directivo

C) ¿Quiénes son los encargados de la evaluación? Es necesario clarificar el rol de los evaluadores y cuál es el nivel de participación de los implicados en su ejecución así como de los receptores.

Responsables de la evaluación	Roles
Profesorado	Ejecutor de la evaluación (responsable en todas sus fases)
Profesor de Psicología y Pedagogía.	Asesor, colaborador en los procesos de evaluación.

D) ¿Qué se tiene que evaluar?

Se han de delimitar los aspectos a evaluar. Concretamente en la Educación emocional los aspectos a evaluar son:

— Los conocimientos básicos sobre emociones
— Las competencias emocionales adquiridas
— Los estados emocionales

E) ¿Cuándo se tiene que evaluar?

Consideramos que son tres los momentos clave en la evaluación de un programa y en cada uno de ellos se pueden distinguir los contenidos de evaluación a tener en cuenta:

Momentos	Contenidos
ANTES	Ev. de necesidades y ev. de diseño
DURANTE	Ev. de proceso y de producto
DESPUÉS	Ev. de proceso y de producto

F) ¿Cuáles son los criterios de evaluación?

Estos deben extraerse de los objetivos propuestos. A partir de aquí se señalarán los indicadores y se concretarán los estándares.

Segunda fase: *Instrumentación y recogida de la información*

G) ¿Qué conviene tener en cuenta para la selección de los instrumentos?

— Que midan aquellas variables y/o dimensiones que se pretenden medir, teniendo en cuenta los aspectos a evaluar: conocimientos, competencias y estados emocionales.

— Que sea lo más simple posible en cuanto al contenido, lenguaje y aplicación.
— Que se ajuste a los requisitos de validez y fiabilidad.

H) ¿Qué tipo de instrumentos y recursos podemos utilizar?

— Cuestionarios *ad hoc*
— Cuestionarios comercializados
— Cuestionarios de carácter cualitativo
— Entrevistas
— Escalas de actitudes
— Autoinformes
— Observación sistemática

Es aconsejable la utilización paralela de diversos instrumentos o técnicas de recogida de información que combinen la obtención de datos cualitativos y cuantitativos.

Tercera fase: *Análisis y valoración de la información*

I) ¿Qué aspectos se habrán de tener en cuenta en el análisis?

— Los criterios establecidos previamente con sus descriptores y stándares.
— La triangulación de la información obtenida por diferentes procedimientos.

J) ¿Qué fases incluye el análisis y valoración de los resultados?

— Podemos distinguir las siguientes fases:

a) Organización y análisis de la información (datos objetivos)
b) Valoración o interpretación de la informacion de acuerdo con los criterios y/o estándares previos.
c) Elaboración de conclusiones, juicios y recomendaciones sobre el programa.

Cuarta fase: *Presentación de los resultados*

K) ¿Cuál es el propósito de la presentación de los resultados?

— Dar a conocer los resultados a las audiencias (conclusiones, o juicios de valor).
— Facilitar la adecuada toma de decisiones sobre el programa mismo o respecto a su futuro

L) ¿Qué aspectos debe tener en cuenta un informe de evaluación?

— Partir del criterio de funcionalidad (propósito de comunicar y transmitir información) adecuando su extensión y redacción a los potenciales receptores.

— Descripción detallada de las intervenciones orientadoras objeto de evaluación y del contexto.
— Descripción de la evaluación efectuada.
— Presentación de los datos y los resultados de evaluación.
— Incluir las conclusiones o juicios de valor que faciliten la oportuna toma de decisiones.
— Conclusiones y recomendaciones

EVALUACIÓN DE NECESIDADES EN EDUCACIÓN EMOCIONAL

La evaluación de necesidades se puede realizar a través de diferentes estrategias y técnicas. A modo ilustrativo presentamos el *Cuestionario de Educación Emocional (CEE)*, elaborado por el GROP, como un instrumento útil tanto para el análisis de necesidades como para la evaluación de la eficacia de la intervención. Este instrumento está pensado para ser cumplimentado por el alumnado y consta de 68 ítems distribuidos en cinco dimensiones: emociones, autoestima, habilidades socio-emocionales, resolución de problemas y habilidades de vida. Se dispone de una versión reducida y autocorregible con las mismas dimensiones pero con sólo 20 ítems (CEE-R). El CEE está diseñado como una escala de tipo Likert en la que los sujetos deben mostrar su acuerdo con las afirmaciones que se presentan marcando una cruz entre las respuestas posibles —nunca, algunas veces, con frecuencia y siempre—. La puntuación de cada ítem va de 0 a 3 puntos respectivamente, excepto en los ítems redactados de forma negativa en los cuales la puntuación es a la inversa. En todas las puntuaciones subtotales y totales del cuestionario se interpretan los valores más altos como un indicador de «buen» estado emocional. En la tabla siguiente se presentan algunos ejemplos de ítems del cuestionario.

Emociones:	— Me vienen ideas a la cabeza que no me dejan dormir por la noche.
Autoestima:	— Me siento a gusto conmigo mismo.
Hab. Socio-Emoc.:	— Estoy satisfecho de las relaciones que tengo con mis familiares.
Res. de Problemas:	— Cuando tengo un problema intento resolverlo.
Hab. de Vida:	— Aprovecho bien mi tiempo libre.

Un análisis estadístico preliminar sobre el CEE puede consultarse en Martínez Olmo *et al.* (2000).

UNA EXPERIENCIA PRÁCTICA

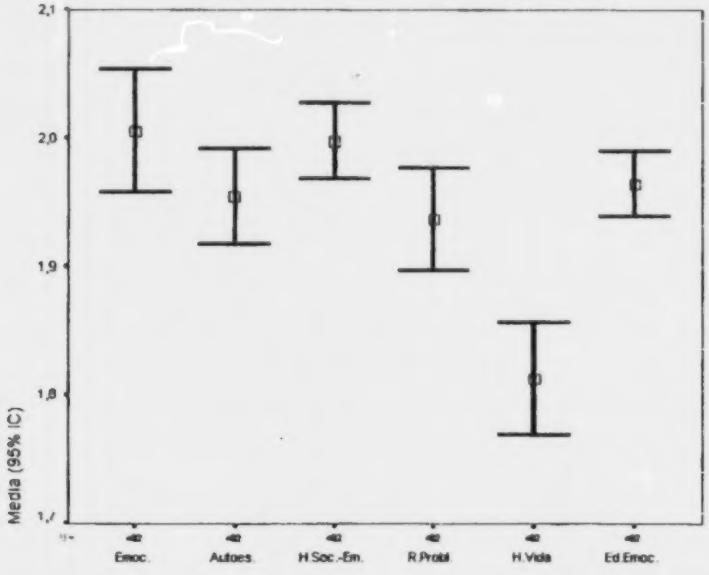
El CEE ha sido aplicado a 8 centros educativos de secundaria con un total de 538 sujetos. Se ha realizado un análisis de datos que incluye descriptivos para cada una de las dimensiones y para el total, se ha confirmado la normalidad de las distribuciones, análisis de ítems, etc. Los datos de la fiabilidad resultante son:

578

Dimensiones (y número de ítems)	alpha de Cronbach
Emociones (11)	0.82
Autoestima (17)	0.79
Hab. Socio-Emocionales (22)	0.69
Resolución de Problemas (14)	0.78
Hab. de Vida (4)	0.20
Educación Emocional (68)	0.87

En la matriz de correlaciones entre las diversas dimensiones del CEE se ha podido observar que son factores muy interrelacionados. Aún así, para facilitar el diseño de programas específicos de actuación y la puesta en práctica de estrategias pedagógicas seguimos pensando que es oportuno establecer una subdivisión que permita clasificar y seleccionar temas a tratar con el alumnado.

No se han detectado diferencias significativas entre las muestras de los distintos centros. La representación gráfica de la media de cada dimensión y del total aparece en la figura siguiente. Los márgenes de cada barra están calculados para un intérvalo de confianza del 95%.



La escasa variabilidad de puntuaciones dentro de cada dimensión hace pensar que un posible valor diagnóstico del instrumento, a nivel individual, consiste en la detección de aquellos sujetos que sobresalen por encima o por debajo de la mayoría. Por otra parte, el diseño del instrumento también se ha concebido como un apoyo y un ele-

mento de contraste a la opinión de los tutores, que son quienes mejor pueden conocer el tipo de información que mide el cuestionario.

En las dimensiones analizadas se han hallado algunas pequeñas diferencias, aunque estadísticamente significativas, en función del sexo y la edad. Así, encontramos que en la dimensión de Emociones las chicas obtienen una media (1,86 puntos) inferior a la de los chicos (2,16), cuya t=6,451 (p=0,000). Es decir, los chicos manifiestan en la muestra mayor estabilidad emocional que las chicas. En esta misma dimensión encontramos diferencias según la edad de los sujetos (F=5,85, p=0,003): el grupo de 12 años obtiene una media superior (2,10) al grupo de 14 años (1,88). Es decir, según este sondeo, a los 12 años se tiene mayor estabilidad emocional que a los 14 años. Esto puede relacionarse con la turbulencia de la adolescencia, donde a los 14 años se está en uno de los momentos más conflictivos.

En la dimensión de autoestima encontramos diferencias según la edad (F=12,668, p=0,000). Las respectivas medias son: 12 años = 2,09; 13 años = 1,96; 14 años = 1,83. Es decir, a medida que aumenta la edad, en el intervalo analizado, tiende a disminuir el nivel de autoestima. La interpretación puede ser anàloga al párrafo anterior.

La edad también se ha mostrado como factor diferenciador en la puntuación global del cuestionario (F=10,203; gl=2; p=0,000). En los 12 años la media es de 2,06; en los 13 es de 1,96; y en los 14 es de 1,90. La interpretación de estos datos puede ser, en gran medida, idéntica a la expuesta en los párrafos anteriores.

Aunque estos datos podrían indicar que es a los 14 años cuando más se manifiesta la necesidad de educación emocional, de acuerdo con el principio de prevención primaria (intervenir antes de que surja el problema), se debería recomendar iniciar los programas en los primeros años de escolaridad.

BIBLIOGRAFÍA

Bisquerra, R., y Álvarez, M. (1998). Los modelos en orientación. En R. Bisquerra *Modelos de orientación e intervención psicopedagógica* (pp. 55-65). Barcelona: Praxis.

Bisquerra, R. (2000). *Educación emocional y bienestar*. Barcelona: Praxis.

Cuadernos de Pedagogía. (1997). *Educación afectiva* (monográfico), n. 261 (septiembre).

Cuadernos de Pedagogía. (1998). *Inteligencia y afectividad* (monográfico), n. 271 (agosto).

DASI, M., Martínez-Vilanova, F., ESIC, R. (1997). Técnicas de negociación. Un método práctico. Manuales prácticos de Marketing.

García , E.M. y Magaz, A. (1998). Test de evaluación de Habilidades cognitivas de Solución de Problemas Interpersonales (EVHACOSPI). Alcalá de Henares: Grupo ALBOR-COHS.

García , E.M. y Magaz, A. (1998). Escala Magallanes de Problemas de Ansiedad (EPANS). Alcalá de Henares: Grupo ALBOR-COHS.

García , E.M. y Magaz, A. (2000). Escala de Actitudes y Valores ante la Interacció Social (ADCAs). Alcalá de Henares: Grupo ALBOR-COHS.

García , E.M. y Magaz, A. (2000). Escalas Magallanes de Hábitos Asertivos en Casa y en el Aula (EMHAS). Alcalá de Henares: Grupo ALBOR-COHS.

580

García, E.M. y Magaz, A. y García Campuzano, R. (1998). Escala Magallanes de Ansiedad (EMANS). Alcalá de Henares: Grupo: ALBOR-COHS.

Goleman, D. (1996). *Inteligencia emocional*, Barcelona: Kairós.

GROP (1998). Educación emocional. En M. Álvarez y R. Bisquerra, *Manual de orientación y tutoría* (pp. 144/27-144/43). Barcelona: Praxis.

GROP (1999). Actividades de educación emocional. En M. Álvarez y R. Bisquerra, *Manual de orientación y tutoría* (pp. 330/27-330/59). Barcelona: Praxis.

Güell Barceló, M. y Muñoz Redon, J. (1999). *Desconócete a ti mismo. Programa de alfabetización emocional*. Barcelona: Paidós.

Güell Barceló, M. y Muñoz Redon, J. (1998). *Desconeix-te a tu mateix. Programa d'alfabetización emocional*. Barcelona: Edicions 62.

Hindle, T. (1998). El estrés bajo mínimos. Biblioteca esencial del ejecutivo.

Martínez Olmo, F., Pérez, N., Fita, E. y Álvarez, M. (2000). *Estudio empírico del Cuestionario de Educación Emocional*. I Congreso Estatal de Educación Emocional, Barcelona.

Sandin, B. (1995). *PANAS* (Versión Española). Madrid: UNED.

Shapiro, L.E. (1998). *La inteligencia emocional de los niños*. Barcelona: Ediciones B.

581

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 601-609

EVALUACIÓN DE PROGRAMAS DE ATENCIÓN A LOS NIÑOS CON DISCAPACIDADES

*María Luisa Dueñas**

Departamento M.I.D.E.
U.N.E.D

RESUMEN

La evaluación de programas de atención a niños con discapacidades ha sido impulsada por los cambios acaecidos en los planteamientos de la educación especial. Estos cambios han contribuido a que se dé un mayor énfasis —del que existía hasta ahora— en que la investigación tenga una orientación aplicada, sea útil y realista. La investigación en este campo debe realizarse en el amplio contexto de las escuelas ordinarias y de los servicios educativos que la sociedad ofrece. Para realizar investigaciones en esta área, los autores suelen coincidir en una serie de dificultades metodológicas relativas a la población, a la representatividad de las muestras, al uso de instrumentos de medida y evaluación, a la generalización de los resultados, etc...

ABSTRACT

The evaluation of programs of attention to handicaped people has been stimulated by the changes in the special education approaches. These changes have contributed to give a more realistic and useful direction to investigation. The research in this field must be made in the wide context of the ordinary schools and the educative services provided by the community. To make these researches, the authors usually agree in some methodological difficulties concerning the population, the representativeness of the research population, measurement instruments and evaluation procedures, the generality of findings, ...

* Facultad de Educación. UNED. Calle Senda del Rey, s/n. 28040-Madrid.
mdueñas@edu.uned.es

582

1. MARCO TEÓRICO

La educación especial es una de las áreas educativas que más se ha desarrollado en los últimos años, sobre todo a partir del Informe Warnock, de 1978, que ha sido el principal impulsor de dicho desarrollo (Wedell, 1985), y en España, a partir de la Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo (LOGSE), dónde se configura un nuevo modelo de educación especial como modalidad educativa integrada en el sistema educativo ordinario.

Este marco legislativo introduce una concepción y una práctica nueva en relación con las necesidades especiales que presentan los alumnos. Ha habido un cambio en los planteamientos: del modelo centrado en el niño y en su diagnóstico diferencial se ha pasado al modelo interactivo de desarrollo del niño. En este modelo, centrado en la interacción entre el niño y los factores ambientales, es preciso tener en cuenta los diferentes entornos en los que el niño vive y aprende, así como la forma en que los cambios de actitudes afectan al desarrollo y al aprendizaje del niño. Es decir, es imposible estudiar y comprender el desarrollo del niño sin tener en cuenta, al mismo tiempo, las numerosas influencias ambientales que le afectan. Este cambio enfatiza la relevancia para la educación ordinaria de conceptos propios de la educación especial porque hace referencia a las ayudas que se deben proporcionar a los niños y jóvenes para su aprendizaje y su desarrollo.

Una consecuencia muy importante de este nuevo enfoque ha sido la necesidad de evaluar la efectividad de las prácticas educativas en el logro de los objetivos (op. cit.), es decir, surgió la necesidad de evaluar la efectividad de los programas de atención a niños con discapacidades.

La evaluación de programas en el campo de la Educación Especial se dirige a evaluar la efectividad de los programas que se establecen en relación con la atención a sujetos con necesidades educativas especiales en orden a una toma de decisiones más racional, normalizadora e integradora. Es preciso reseñar que los modelos de evaluación de programas en educación especial son los mismos que para la educación ordinaria (Mittler, 1985; Garanto, 1989). Precisamente Garanto (1989) propone una serie de estrategias metodológicas no «especiales» que resultan más idóneas que otras para la evaluación de programas de educación especial, a saber:

- Estrategias descriptivas y correlacionales, ya que ofrecen una descripción y explicación de las variables que intervienen en el proceso, son más factibles de utilizar que las causales (experimentales), aunque éstas no tienen porque desecharse. El diseño experimental se presta sobre todo para la realización de fines que quedan fuera del programa inmediato (Weis, 1975).
- Es conveniente aplicar programas «alternativos» al grupo de control y al grupo experimental para evaluar con cual de ellos los objetivos fijados son alcanzados con mayor éxito, de tal forma que pueda ser factible la toma de decisiones sobre la implantación y generalización del mejor.
- La comparación pretest-postest es la más idónea cuando solo existe un único programa. Un único grupo y un único programa es la forma más habitual en la

583

investigación evaluativa, dependiendo en ese caso la validez de los criterios de medida adoptados.

- La metodología N = 1 en educación especial es muy adecuada. Los diseños de sujeto único (caso único, intensivo, intra-sujeto, etc.) se aplican en grupos de sujetos o individuos en tanto en cuanto éstos son considerados como una unidad.

2. ASPECTOS METODOLÓGICOS

La dimensión metodológica constituye, junto con la dimensión conceptual-teórica, la componente clave de la evaluación de programas (Colás, 1993). Los procesos metodológicos para llevar a cabo la evaluación de programas cubren una extensa gama de posibilidades.

Durante los últimos años, los cambios ocurridos en los planteamientos de la educación especial (Dueñas, 1994) han contribuido a un cambio en el enfoque metodológico, en la orientación general que se está produciendo en la investigación educativa en dicho campo. En particular, se da un mayor énfasis del que existía hasta ahora, en que la investigación tenga una orientación aplicada: es necesario evaluar los programas y los procesos concretos que intervienen en la educación especial. La investigación debe ser útil y realista, y debe analizar los problemas en toda su complejidad.

En muchos procesos de evaluación de programas de educación especial se han encontrado improvisaciones y errores metodológicos (Tindal, 1985), lo que puede ser debido al afán por presentar «hechos» sin preocuparse de su calidad.

La evaluación de un programa de educación especial, puesto que es decisiva para una toma de decisiones, tiene que realizarse con seriedad y rigor metodológico, reconociéndose como un proceso verdaderamente complejo (Garanto, 1989).

El diseño, la implantación y la evaluación de un programa de intervención educativa controlando los logros de los alumnos que lo siguen, es una tarea compleja y difícil, y más aún cuando va dirigido a los alumnos con necesidades educativas especiales.

Para realizar investigaciones en este campo, los autores suelen coincidir en una serie de dificultades metodológicas (De Miguel, 1986; Schindele, 1985; Dueñas, 1994):

a) Variabilidad de la población

La población suele presentar las características de ser extremadamente heterogénea y presentar una gran variabilidad dentro de los grupos. Existe una gran complejidad a nivel individual causada por la discapacidad y el tratamiento específico, lo que suele considerarse como la mayor dificultad de las investigaciones en educación especial (Cronbach y Snow, 1979) y solo puede evitarse a través de los diseños que minimizan las diferencias individuales.

b) Problemas de muestras

Las dificultades en este campo son serias y muy distintas: problemas de elección de sujetos al azar (aleatorización) y dificultades para obtener muestras representativas, tanto en lo que se refiere al número como a una determinada discapacidad y para la constitución de un adecuado grupo de sujetos espe-

cialmente en los diseños experimentales referentes a grupos (Bryk y Weisber, 1977).

c) Imprecisión en la operativización e instrumentalización de las variables

Cada investigador, al efectuar los diseños, no solo establece una definición «peculiar» de los «handicaps», sino que parte de unos criterios e instrumentos de medida que difícilmente son transferibles de una situación a otra. Además, es preciso tener en cuenta el cambio que se ha producido en los criterios para la asignación de niños con discapacidad psíquica, el sistema de clasificación no está ya basado solo en el C.I. sino en el tipo de intensidad de los apoyos que requieren las personas con retraso mental.

d) Uso de instrumentos de medida y evaluación

Es necesario cuidar al máximo la elección y posterior aplicación de las pruebas en el sentido de que éstas deben ser adecuadas a la población a la que se dirige el programa (Salvia y Ysseldyke, 1981; Dueñas, 1990), ya que a veces los problemas surgen ante la necesidad de adaptar tests y procedimientos de evaluación a la discapacidad especifica del sujeto, debido a la falta de instrumentos normativos para la valoración de niños con discapacidades y a la multiplicidad de programas de intervención disponibles (Ulrey y Rogers, 1982).

Hay que definir cuidadosamente el objetivo exacto de tal evaluación y en función de ello establecer consideraciones detalladas sobre los tests existentes, establecimiento de criterios de puntuación, etc...., pero en todo ello se está muy por debajo de los estándares de tests válidos (Hogg y Raynes, 1987).

e) Ausencia de controles apropiados

Si se tienen presentes las causas o fuentes de invalidez, peligro de errores en la validez interna considerando las variables de intervención y su interacción o los efectos de la regresión (Wolf, 1979), fácilmente se deduce que los trabajos en este campo son muy vulnerables a las fuentes de invalidez. Por todo ello, a veces se confunde lo que es una variación normal de los fenómenos y los efectos de la intervención y /o situación.

f) Dificultades en la generalización de los resultados

Los problemas anteriores relativizan la posibilidad de generalizar los resultados. Sólo cuando los diseños se han efectuado con una tecnología ampliamente implementada y demostrada empíricamente, cabe pensar en la generalización. Por el propio contexto, gran variabilidad, en el que se producen este tipo de investigaciones y las dificultades para replicar los experimentos, esta situación es ciertamente poco frecuente (Brennan, 1984).

g) La propia naturaleza de la educación especial

El proceso mismo de educación especial, por su naturaleza dinámica, por su complejidad y multidimensionalidad, por su singularidad e individualidad, puede

tener una gran influencia en la investigación de la evaluación de la intervención (Mc. Reynolds y Kearns, 1982). Las estrategias y métodos de investigación en este área pueden estar también delimitadas por el contexto institucional en el que la investigación se realiza y por las consideraciones éticas y morales que es necesario tener en consideración cuando la investigación concierne a personas con condiciones excepcionales en su desarrollo y aprendizaje (Schindele, 1985).

3. DIMENSIONES DEL PROGRAMA A EVALUAR

Para nosotros, la evaluación constituye un proceso continuo que tiene lugar en los distintos momentos del programa: planificación, ejecución y resultados. En dicho proceso, los objetivos variarán según sea el referente.

Como dimensiones o contenidos básicos de un programa de atención a niños con discapacidades suelen señalarse los siguientes: identificación de necesidades; diseño del programa y/o programación de objetivos; implantación y ejecución del programa; y, evaluación del programa a través de los resultados (Lieberman y Mc Neil, 1982; Caride, 1989; Dueñas, 1994).

La identificación de necesidades es fundamental ya que define el entorno o marco en el que se va aplicar el programa, detecta la naturaleza de las necesidades especiales de los alumnos a los que se dirige el programa e identifica carencias apreciables en el contexto que pueden poner en peligro la consecución de los objetivos.

El análisis de necesidades ocurre durante la fase de planificación de un programa, antes de que se haya ofertado. La evaluación de necesidades constituirá el punto de partida para identificar situaciones deficitarias e insatisfactorias que reclamen solución; es un proceso complejo a través del cual se deciden, teniendo en cuenta los recursos, las prioridades de actuación (Tejedor, 1990). Esta fase sirve de base a las decisiones de planificación para definir sus objetivos.

El diseño del programa y/o la planificación de objetivos, debe ser adecuado a las necesidades detectadas, y prever los recursos materiales y humanos —staff de profesionales— necesarios para la consecución de las metas y los objetivos propuestos.

Los objetivos, en el área que nos ocupa, pueden ser muy diversos: constatar los logros obtenidos por los alumnos sometidos a un programa de intervención educativa; conocer cuál es la percepción que tienen las personas que participan en un determinado programa; recoger información para fomentar la reflexión critica de los procesos para tomar decisiones adecuadas a cada situación específica por parte de las personas implicadas en el proceso.

Implantación y ejecución del programa: El desarrollo del programa puede encontrar dificultades no previstas o descubrir vías alternativas consideradas más adecuadas al nivel de conocimiento de la realidad; por ello, la evaluación de proceso se convierte en una dimensión especialmente útil en un ámbito donde con gran frecuencia es más relevante el cómo que el cuánto o el qué, sobre todo cuando se trate de mejorar estos dos últimos aspectos.

Evaluación del programa o de resultados: La evaluación de producto tiene por finalidad conocer hasta que punto el programa ha satisfecho las necesidades del grupo

586

al que pretendía servir utilizando una combinación amplia de técnicas para obtener una visión global de los resultados del programa. A nivel individual, un resultado puede considerarse como de éxito, o fracaso, si el programa ha satisfecho, o no, una necesidad previamente diagnosticada (Stuflebeam et al., 1987). Este tipo de evaluación informa acerca de la calidad y eficacia del programa.

En síntesis, todas estas dimensiones nos proporcionan la siguiente información: necesidades que existen y hasta que punto los objetivos propuestos satisfacen las necesidades; descripción del programa de intervención y análisis conceptual de la adecuación de la propuesta elegida a los objetivos; grado de realización del plan de intervención propuesto y descripción de las modificaciones; resultados obtenidos y grado en que se han satisfecho las necesidades.

4. CONSIDERACIONES FINALES

En consonancia con los nuevos planteamientos de la educación especial anteriormente reseñados, la investigación en este campo debe realizarse en el amplio contexto de las escuelas ordinarias y de los servicios educativos que la sociedad ofrece (Mittler, 1985).

Determinar qué tipo de programas de intervención o prácticas educativas son más eficaces para satisfacer las necesidades, tanto escolares como sociales, en clases ordinarias constituye una de las preocupaciones más importantes en este área (Slavin y Madden, 1986).

La evaluación de programas en educación especial debe plantearse desde una perspectiva amplia e integral, analizando previamente la entidad de las metas perseguidas, dándose coherencia entre las metas del programa y demandas / expectativas / necesidades, teniendo en consideración las características del contexto personal/social en el que el programa se desarrollará. A esta visión, el modelo CIPP de Stuflebeam ha dado, a nuestro juicio, una respuesta notablemente adecuada, por ser el modelo en el que aparece explícitamente la orientación hacia la toma de decisiones, proporcionando información útil para la misma.

Se debe concretar de qué programa se trata en cada caso, cuál es el curriculum aplicado, qué roles desempeñan los distintos profesionales, qué procedimientos y técnicas se aplican, cuáles son las actitudes de los profesionales y de la comunidad educativa. Las conclusiones de la evaluación deberán restringirse a ese programa y ese contexto, pudiendo generalizar los resultados con certeza sólo cuando la replicación de los mismos y su comparación con otros programas resulten positivos sistemáticamente (Verdugo, 1989).

A pesar de la diversidad tanto conceptual/teórica como metodológica que existe en la evaluación de programas de atención a niños con discapacidades, todos ellos participan de un objetivo común: generar conocimiento científico-evaluativo a través de una metodología que intenta ser científica y que adopta distintas fórmulas pero que comparte criterios comunes tales como racionalidad, contrastabilidad, sistematización, carácter práctico, etc... (Colás, 1993).

587

Finalmente, y siguiendo en parte a Schindele (1985), proponemos unas orientaciones metodológicas que pueden servir de guía para realizar una mejor investigación en este campo:

1. *Aplicación de diseños de investigación más adecuados*
   La aplicación de diseños inadecuados es con frecuencia la causa de que los resultados sean poco relevantes y significativos. Es preciso aumentar la aplicación de diseños experimentales de caso único, series temporales, diseños intrasujeto (Whitehead, 1982; Mc Reynolds y Kernan, 1982; De Miguel, 1986; Dueñas, 1990). Estos diseños parecen en la mayoría de los casos particularmente relevantes para temas específicos y cuestiones de investigación en esta área para afrontar la complejidad de los temas investigados y la multiplicidad de los objetivos correspondientes.

2. *Utilización de la metodología de investigación tanto cuantitativa como cualitativa*
   Cuando los problemas de la investigación son complejos son necesarios diseños de investigación mixta (Dueñas, 1990), que parecen especialmente apropiados a proyectos en los que ambas metodologías son necesarias y donde la recogida y evaluación de datos están dirigidas tanto a los grupos como a la investigación intensiva de situaciones individuales. Murphy y Bryan (1980) presentan la combinación de diseños múltiples como una práctica alternativa a la valoración y evaluación en educación especial.

3. *Aplicación más adecuada y más sistemática de la metodología cualitativa.*
   La importancia de la aproximación cualitativa (etnográfica, ecológica, investigación-acción, ...) en educación especial ha sido subrayada por un gran número de especialistas. Este tipo de investigaciones se preocupa más por los procesos que por los resultados y tiende a analizar los datos inductivamente (Schindele, 1985).

4. *Aplicación creciente de los principios de la investigación-acción.*
   La investigación-acción en este campo parece muy adecuada para la evaluación de métodos instructivos, materiales y programas (Wolery y Harris, 1982). Esta aproximación implica una óptima posibilidad para la aplicación de los resultados de investigación porque elimina la distancia entre el investigador y el profesor.

En síntesis, la metodología de investigación en el área que nos ocupa está necesitada, en gran medida, de un desarrollo más sistemático, tanto en aproximaciones y diseños como en evaluaciones estadísticas. Muchas investigaciones son aún poco válidas, tienen lagunas en su significado y escasa relevancia en la práctica educativa. Deberían promoverse investigaciones interdisciplinares y discusiones/debates sobre los objetivos y los métodos de las investigación y sobre su impacto tanto sobre los sujetos con necesidades educativas especiales como sobre las prácticas educativas.

588

BIBLIOGRAFÍA

Bogdan, R.C. y Biklen, S.K. (1982). *Qualitative Research for Education: An Introduction to Teory and Practice*. Hemel Hempstead: Allwyn and Bacon.

Bryk, A.S. y Weisberg, H.I. (1977). Use of the non equivalent control group design when subjects are growing. *Psychological Bulletin*, 84, 5, pp. 950-962.

CARIDE, (1989). «De la evaluación de necesidades a la evaluación de programas sociales en el desarrollo comunitario», cit. por F.J. Tejedor (1990).

Colas, M.P. (1993). «Concepto, funciones y etapas en la evaluación de programas» en pp. 17-31 de M.P. Colas y M.A. Rebollo: *Evaluación de Programas. Una guía práctica*. Sevilla: Kronos.

Cronbach, L. y Snow, R. (1979). *Aptitudes and instructional methods: a handbook for research on interactions*. New York: Irvinston.

De Miguel, A. (1986). «Líneas de investigación e Educación Especial» en pp. 62-85 de S. Molina (Dir.): *Enciclopedia Temática de Educación Especial (tomo 1)*. Madrid: CEPE.

Dueñas, M.L. (1990). Estudio empírico de las variables que inciden en el progreso de los alumnos deficientes mentales integrados. *Revista de Investigación Educativa*, vol. 8, no. 15, pp. 19-33.

Dueñas, M.L. (1994). *Métodos de diagnóstico e intervención educativa en la deficiencia mental*. Madrid: UNED (col. Cuadernos nº 135).

Garanto, J. (1989). «Modelos de evaluación de programas educativos», en pp. 43-78 de M.C. Abarca (Coor.): *Evaluación de Programas Educativos*. Madrid: AEDES. Edit. Escuela Española.

Hog, J. y Raynes, N. (Eds) (1987). *Assessment in Mental Handicap*. Cambridge (Mas.): Brookline Books.

Lieberman, L.M. y Mc Neil, D.M. (1982). Evaluating Special Education Programs. *Journal of Learning Disabilities*, 15 (2), pp. 121-122.

McReynolds, L. y Kearns, K.P. (1982). *Single Subject Desings for Intervention Research in Comunicative Disorders*. Lancaster: MTP Press.

Mittler, P. (1985). «Approaches to Evalution in Special Education: Concluding Reflections» en pp. 165-176 de S. Hegarty y P. Evans (Eds): *Research and Evaluation. Methods in Special Education*. Windsor: NFER-NELSON.

Murphy, R. y Bryan, A. (1980). *Multiple-baseline and multiple-probe designs: Practical alternatives for especial education assessment and evaluation*. The Journal of Special Education, 14, 3, pp. 325-335.

Salvia, J. e Ysseldyke, J. (1981). *Assessment in Special and Remedial Education*. Boston: Houghton Mifflin Campany.

Schindele, R.A. (1985). Research Methodology in Special Education: A Framework Approach to Special Problems and Solutions, en pp. 3-24 de S. Hegarty y P. Evans (Eds). *Research and Evaluation. Methods in Special Education*. Windsor: NFER-NELSON.

Slavin, R.E. y Madden, N.A. (1983). La integración en las clases ordinarias de los alumnos con retraso escolar. *Perspectivas*, 4, pp. 473-493 (original en *Review of Educational Research*, 53 A , pp. 519-569).

589

Stuflebeam, D.L. y Shinkfield, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica.* Barcelona: Paidos-MEC.

Tejedor, F.J. (1990). Perspectiva metodológica del diagnóstico y evaluación de necesidades en el ámbito educativo. *Revista de Investigación Educativa*, vol. 8, nº 16, pp. 15-37.

Ulrey, G. y Rogers, S.J. (1982). *Psychological Assessment of Handicapped Infants and Young Children*. New Yok: Thiene-Stratton.

Whitehead, J. (1982). *The Design and Analysis of Sequential Clinical Trials*. Chichester: John Wëley.

Wolery, J. y Harris, S.R. (1982). Interpreting results of single-subject research designs. *Physical Terapy*, 62, 4, pp. 445-452.

Wolf, R.L. (1979). *Some generic considerations in soio-educational evaluation*, cit. en R.A. Schindele (1985).

590

TERCERA PARTE

Planteamientos que vienen siendo utilizados en la realización de evaluaciones en los Departamentos del área de MIDE

PRESENTACIÓN

Como se indicó en la presentación del presente volumen, el número monográfico se encuentra estructurado en tres grandes partes. Las dos anteriores fueron encargadas a diferentes profesores del área con la pretensión de que ofrecieran su visión personal bien sobre *planteamientos metodológicos generales,* en la primera parte, bien sobre *planteamientos metodológicos especialmente adecuados a la evaluación de cierto tipo de programas,* en la segunda.

La tercera parte, que ahora comienza, se concibió como un medio para ofrecer al lector el *mapa metodológico del área* en lo relativo a la evaluación de programas. Con la intención de que presentara una cierta unidad, se elaboró un esquema que sirviera de referencia a la hora de que los diversos departamentos presentaran su aportación, aunque dejando abierta la necesaria flexibilidad para evitar cualquier rigidez.

Si bien en el documento remitido a los departamentos se indicaba, de modo expreso, que «No se trata de informes de investigación sino de trabajos de síntesis, realizados por una persona encargada de ello en cada Departamento», las respuestas que encontrará el lector ponen de relieve una gran diversidad, algo que si por un lado, pone de relieve la riqueza del área, aunque, por otro probablemente hará más difícil para el lector alcanzar la representación del *mapa metodológico del área.*

A continuación se presentan las aportaciones realizadas por profesores de las universidades de Alicante, Barcelona, Cádiz, Complutense, Granada, Jaén, Málaga, Murcia, Navarra, País Vasco, Salamanca, Santiago de Compostela, Sevilla y UNED.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 615-620

UNIVERSIDAD DE ALICANTE

María Cristina Cardona Moltó[1]
Departamento de Psicología de la Salud

TÍTULO DEL PROGRAMA: **ESTRATEGIAS PARA LA MEJORA Y ADAPTACIÓN DE PROGRAMAS INSTRUCTIVOS**[2]

JUSTIFÍCACIÓN DEL PROGRAMA

Una tendencia cada vez más arraigada en el campo de la educación es la inclusión de los alunmos con necesidades educativas especiales en las aulas ordinarias. Esta tendencia se basa en el supuesto de que el aula ordinaria es el lugar idóneo para la educación de la mayoría de los estudiantes, tanto de los más como de los menos capaces, por su poder para potenciar su desarrollo personal, académico y social. Los recientes cambios demográficos en la población escolar han hecho aumentar, por otro lado, el número de estudiantes con riesgo de fracaso escolar, con lo que la diversidad y heterogeneidad en las aulas se ha convertido en la norma, más que en la excepción. Este estado de cosas, se ha ido extendiendo progresivamente hacia niveles superiores de enseñanza, en la medida en que ésta se ha ido generalizando. Por tanto, en relativamente pocos años, las condiciones de trabajo de los profesores han cambiado y, con ello, sus responsabilidades docentes han ido en aumento.

La mayoría de los expertos en el campo están de acuerdo en que para hacer frente a las nuevas demandas son necesarios cambios profundos. Cambios en el concep-

---

1 Miembros del equipo de investigación: María R. Bermejo García, Rosa Mª Carda Ros, Pedro F. Gómez Canet, Juan P. González Gómez, María C. González Gómez, María E. González Sánchez, María D. Jiménez Alegre, Ana Lattur Devesa, Antonio Vallés Arándiga.

2 Este programa forma parte del proyecto de investigación «Estrategias y Prácticas Inclusivas de Adaptación de la Enseñanza en Centros Educativos de la Provincia de Alicante», referencia GV99-56-1-07, subvencionado por la Conselleria de Cultura, Educación y Ciencia de la Generalitat Valenciana dentro del programa Proyectos de Investigación Científica y Desarrollo Tecnológico 1 + D «Generalitat Valenciana».

to mismo de educación y cambios en el papel y la función de la escuela y de sus profesores (Ferguson, 1997). El cambio operado en poco tiempo en las escuelas afecta tanto al *qué* como al *cómo* los profesores enseñan (Delpit, 1996). Sin embargo, mientras las instituciones educativas han cambiado, no lo han hecho al mismo ritmo las prácticas instructivas de los profesores, que en muchos aspectos permanecen obsoletas. Los cambios, dicen los expertos, van a depender de la voluntad y de la habilidad de los profesores para alterar sus creencias y percepciones, y también sus prácticas. La literatura de investigación sugiere que ésta no es una tarea fácil y que el profesorado muestra una cierta resistencia al cambio. Esta resistencia se atribuye a inseguridad y/o falta de competencia profesional (Schumip y Vaughn, 1991; Semmel, Abemathy, Butera y Lesar (1991), a la insuficiente disponibilidad de apoyos (Wholery *et al.*, 1995) o simplemente a la opinión de que los cambios instructivos, aunque deseables, no son posibles en la práctica (Minke, Bear, Deemer, Griffin, 1996; Scott, Vitale y Masten, 1998).

Este equipo de investigación arguye que los cambios no pueden venir dados por inercia o simple imposición, sino más bien como resultado de las acciones normativas que se emprendan. El profesorado necesita no sólo mejorar su formación, sino además que se le ofrezcan oportunidades para experimentar por sí mismo situaciones de éxito y experiencias de aprendizaje desde su puesto de trabajo. Necesita, en definitiva, el soporte necesario para un desarrollo profesional sostenido con experiencias de colaboración entre universidad y centros educativos (Blanton, Griffin, Winn y Pugach, 1997). Así, este programa surge con la pretensión de contribuir al desarrollo profesional del profesorado de los niveles obligatorios de la enseñanza desde sus propios centros, a través de su participación en un programa de formación en estrategias de mejora y adaptación de los programas instructivos. Sus objetivos básicos son: (1) explorar las creencias y percepciones del profesorado en tomo a una pedagogía inclusivo integradora de las diferencias del alumnado en el aula, y analizar cómo los profesores interpretan y transforman sus creencias como resultado de su participación en el programa; (2) identificar el tipo de adaptaciones de la enseñanza que hacen los profesores de educación infantil, primaria y secundaria y cómo las juzgan; (3) enseñar al profesorado procedimientos para adaptar los programas, y valorar la eficacia del uso separado y combinado de estrategias (generales e individuales) de adaptación de la enseñanza; (4) dotarle de procedimientos y técnicas para diseñar, implementar y evaluar adaptaciones específicas para aquellos alumnos con muy bajo o muy alto rendimiento para quienes las estrategias de adaptación general pudieran resultar no satisfactorias; (5) por último, presentar a la administración educativa autonómica propuestas de mejora en el sistema de formación y desarrollo profesional docente y estrategias de atención a la diversidad en los centros viables y eficaces.

El programa tiene una función formativa, sumativa y también socio-política, como se desprende de sus objetivos. Se encuadra en un paradigma ecléctico por su uso combinado de procedimientos metodológicos cuantitativos y cualitativos, y se inspira en un modelo proceso-producto que va más allá de los meros resultados con la pretensión de que éstos influyan en la toma de decisiones futuras.

594

METODOLOGÍA DE EVALUACIÓN

En el curso académico entrante 2000-2001, inicia el programa su segunda fase de desarrollo. En esta fase se hallan involucrados 25 centros públicos de educación primaria de la provincia de Alicante y un total aproximado de 75 profesores voluntarios (tres por centro: 1 profesor que implementará el programa, 1 profesor emparejado por nivel que servirá de control, un profesor de pedagogía terapéutica) y 1.500 alumnos, de ellos, casi un centenar, son alumnos con necesidades educativas especiales. El programa en su segunda fase se caracteriza por (a) la implementación de adaptaciones generales en la enseñanza de una materia fundamental del currículum, (b) su seguimiento, y (c) en función de la evaluación resultante, el reajuste del programa y/o la introducción de adaptaciones individuales (eligiendo 2 alumnos por clase, uno de muy alto y otro de muy bajo rendimiento, con sus respectivos controles). Los elementos o componentes esenciales que guían el proceso son: (1) la evaluación basada en el currículum (EBC), (2) el uso de una metodología de las lecciones altamente estructurado (15′ instrucción directa, 25 minutos trabajo en parejas o pequeño grupo, 10′ control), (3) la evaluación continua y normativa, (4) el soporte al profesor en el aula. El programa incluye un proceso de formación del profesorado participante en la metodología y las técnicas que exige la implementación del programa que se llevará a cabo a través de seminarios de trabajo (uno por trimestre) en la universidad y de la constitución de grupos de trabajo en cada centro de profesores (CEFIRE) de la provincia, coordinados por un profesor de la universidad. Los pormenores del programa se presentan con mayor detalle en la Tabla I.

INFORME

Los destinatarios o personas interesadas en conocer el grado de consecución de las metas y los resultados obtenidos en el programa son los profesores generalistas y especialistas de los niveles obligatorios de la enseñanza, administradores educativos y personal de apoyo (inspectores, servicios psicopedagógicos, asesores de los CEFIRES), así como los responsables de las DGs de Ordenación e Innovación Educativa y Formación del Profesorado de la Conselleria de Cultura, Educació i Ciència de la Comunidad Valenciana. Aunque la cobertura del programa es sectorial (nivel provincial), sus características y datos más relevantes se diseminarán a través de:

- Jornadas y talleres de trabajo en los CEFIREs de la provincia y de la Comunidad Valenciana.
- Creación de grupos y seminarios permanentes de trabajo.
- Elaboración y difusión de informes técnicos.
- Ponencias y comunicaciones en congresos de ámbito nacional e internacional (AIDIPE, EERA, AERA, etc.).
- Publicación de manuscritos en revistas de alto impacto nacional e internacional (Revista de Investigación Educativa, American Educational Research Journal, Journal of Teacher Education, etc.).

Además, paralelamente al curso de desarrollo del programa, se elaborarán materiales de soporte como vídeos, guías didácticas y material docente y para la evaluación diferenciado por materias. De momento, los productos generados en la fase pre-experimental del programa han sido la elaboración de diversos cuestionarios y escalas diseñadas ad hoc que, junto con los primeros informes, se darán a conocer en breve.

BIBLIOGRAFÍA

Blanton, Griffin, Winn y Pugach (1997). *Teacher education in transition: Collaborative programs to prepare general and special educators*. Denver: Love Publishing.

Cardona, M.C. (2000). *Escala de Adaptaciones de la Enseñanza* (ESAE). Formas General y Específica. Universidad de Alicante, Dpto. Psicología de la Salud: Alicante.

Cardona, M.C., Gómez, P.F. y Gónzalez Sánchez, M.E. (2000). *Cuestionario de Percepciones del Profesor acerca de la Pedagogía Inclusiva*. Universidad de Alicante, Dpto. Psicología de la Salud: Alicante.

Cardona, M.C., Lattur, A. y Gónzalez Gómez, J.P. (2000). *Cuestionario de Datos Generales del Profesor y del Centro*. Universidad de Alicante, Dpto. Psicología de la Salud: Alicante.

Delpit, L.D. (1996). *Other people's children: Cultural conflict in the classroom*. New York: New Press.

Ferguson, D.L. (1997, marzo). *Changing tactics: Research embedding inclusion reforms within general education restructuring efforts*. Trabajo presentado a la reunión anual de la American Educational Research Association. Chicago, IL.

Minke, K.M., Bear, G.G., Deemer, S.A. y Griffin, S.M. (1996). Teachers'experiences with inclusive classrooms: Implications for special education reform. *The Journal of Special Education*, 30(2), 152-186.

Schumm, J.S. y Vaughn, S. (1991). Making adaptations for mainstreamed students: General classroom teachers'perspectives. *Remedial and Special Education*, 12(4), 18-27.

Scott, B.J., Vitale, M.R. y Masten, W.G. (1998). Implementing instructional adaptations for students with disabilities in inclusive classrooms. A literatura review. *Remedial and Special Education*, 19(2), 106-119.

Semmel, M.L., Abernathy, T.V., Butera, G. y Lesar, S. (1991). Teacher perceptions of Regular Education Initiative. *Exceptional Children*, 58, 9-24.

Wholery, M., Werts, M.G., Caldwell, N.K., Snyder, E.D. y Lisowski, L. (1995). Experienced teachers' perceptions of resources and supports for inclusion. *Education and Training in Mental Retardation and Developmental Disabilities, March*, 15-26.

NOTA:

Este equipo de investigación quiere agradecer a los CEFIREs de Alcoy, Alicante, Benidorm, Elche, Elda y Orihuela su inestimable colaboración en el proceso de administración de los instrumentos DGPC., ESAE y PEPPI.

TABLA 1

METODOLOGÍA DE LA EVALUACIÓN DEL PROGRAMA

	FASE PRE-EXPERIMENTAL	FASE EXPERIMENTAL		
		1er Período	2º Período	3er Período
Período	Enero 1999-julio 2000	Septiembre-diciembre 2000	Enero-marzo 2001	Abril-junio 2001
Objetivo/s	- Explorar las creencias y percepciones del profesorado en torno a la diversidad. - Identificar el tipo de adaptaciones de la enseñanza que realizan los profesores y si éstas difieren en función de la etapa educativa.	- Examinar las creencias y percepciones de una muestra representativa de profesores de la provincia (unos 400 profesores). - Diseñar, implementar y evaluar estrategias de adaptación general de los programas de una asignatura a través de la EBC.	- Examinar la evolución de las creencias y percepciones de profesores informativos que implementan el programa. - Continuar con el diseño, implementación y evaluación de estrategias de adaptación general. - Diseñar, implementar y evaluar adaptaciones individuales para los alumnos más avanzados/retrasados.	- Realizar un seguimiento de los objetivos del período anterior.
Actividad	- Diseño de instrumentos. - Pilotaje (participantes: un centenar de profesores de todos los centros públicos de una misma localidad alicantina).	- Describir el pensamiento del profesor en torno a la diversidad en una muestra de 75 profesores voluntarios (3 prof x 25 centros). - Valorar la eficacia de las estrategias generales de adaptación sobre el rendimiento de los alumnos en función del tipo de agrupamiento (pequeño grupo vs compañero tutor) y el nivel de conocimiento en la materia (bajo, medio, alto).	- Describir e interpretar las percepciones del profesorado. - Valorar la eficacia de las estrategias generales e individuales diseñadas. - Comparar la viabilidad y eficacia: adaptación general vs individual.	- Observar la evolución de las percepciones del profesorado. - Valorar la eficacia de las estrategias generales e individuales diseñadas. - Comparar la viabilidad y eficacia: adaptación general vs individual.
Diseño	No-experimental, descriptivo	No-experimental, descriptivo Experimental, factorial 2 x 3	No-experimental, descriptivo Experimental, factorial 2 x 3 Diseños de línea base múltiple (n=1)	No-experimental, descriptivo Experimental, factorial 2 x 3 Diseños de línea base múltiple (n=1)
Técnicas recogida información	- Cuestionario Datos Generales del Profesor y del Centro (Cardona, Lattur, Gónzalez Gómez, 2000). - Cuestionario de Percepciones del Profesor acerca de la Pedagogía Inclusiva (Cardona, Gómez y González Sánchez, 2000). - Escala de Adaptaciones de la Enseñanza (ESAE). Formas General y Específica (Cardona, 2000).	- Cuestionarios DGPC, PEPPI y ESAE. - Observación participante y entrevista. - Diarios del profesor, planes semanales de la materia. - Test referidos a criterio. - PROLEC, BAGYG y otras pruebas de aptitud. - Portfolios.	- Observación participante y entrevista. - Diarios del profesor. - Planes semanales de la materia. - Programas individuales. - Test referidos a criterio. - Portfolios.	- Observación participante y entrevista. - Diarios del profesor. - Planes semanales de la materia. - Programas individuales. - Test referidos a criterio. - Portfolios.

TABLA 1 (continuación)
METODOLOGÍA DE LA EVALUACIÓN DEL PROGRAMA

	FASE PRE-EXPERIMENTAL	FASE EXPERIMENTAL		
		1er Período	2° Período	3er Período
Criterios y referentes	- Psicométricos (adecuación, fiabilidad, validez). - Relevancia, comparabilidad y significatividad de la información recogida.	- Cambios y evoluciones en las percepciones del profesorado. - Viabilidad, credibilidad, eficacia. - Referencia criterial y progreso sobre el propio programa.	- Mantenimiento/transformación de las percepciones. - Viabilidad, credibilidad, eficacia. - Referencia criterial y progreso sobre los propios programas.	- Mantenimiento/transformación de las percepciones. - Viabilidad, credibilidad, eficacia. - Referencia criterial y progreso sobre los propios programas.
Técnicas de análisis de datos	- Análisis de escalas. - Análisis factorial. - Descriptivos, porcentajes. - Comparación medias por etapa educativa.	- Análisis de contenido de las reflexiones y comentarios de los profesores. - Descriptivos y porcentajes de R en los tests. - MANOVA y MANCOVA.	- Análisis de contenido de las reflexiones y comentarios de los profesores. - Descriptivos y porcentajes de R en los tests. - MANOVA y MANCOVA. - Inspección de gráficos y análisis de tendencias en diseños n = 1. . .	- Análisis de contenido de las reflexiones y comentarios de los profesores. - Descriptivos y porcentajes de R en los tests. - MANOVA y MANCOVA. - Inspección de gráficos y análisis de tendencias en diseños n = 1.
Tipo de decisiones	- Introducir mejoras y perfeccionar los instrumentos. - Identificar formas de pensar y actuar del profesorado. - Detectar necesidades de formación.	- Ajuste continuo de los programas de la asignatura como resultado de la evaluación. - Identificar a 2 alumnos, uno con rendimiento muy alto y otro muy bajo, y diseñar las correspondien[illegible] adeptaciones individuales.	- Introducir de modificaciones en los programas generales e individuales, si procede.	- En función del progreso, extender el programa al profesorado de ESO. - Probar la viabilidad y eficacia del programa con otras materias y/o asignaturas en educación primaria. - Elevar a la administración propuestas y recomendaciones en el sistema de formación del profesorado y ordenación académica en los centros.

598

BEST COPY AVAILABLE

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 621-626

UNIVERSIDAD DE BARCELONA

Flor Cabrera (coord.), Jaume del Campo, Julia V. Espín, M. Ángeles Marín, Mercedes Rodríguez y M. Paz Sandin

TÍTULO DEL PROGRAMA: **EVALUACIÓN DEL PROYECTO EPIKOUROS DE INSERCIÓN SOCIOLABORAL DE INMIGRANTES**

OBJETO EVALUADO

El proyecto Epikouros del Ayuntamiento de Barcelona representa un modelo de servicio específico de orientación para la inserción sociolaboral dirigido a personas procedentes de la inmigración exterior y con situación socioeconómica deficitaria. El proyecto pretendía responder a las necesidad peculiares que estas personas tienen en su proceso de inserción sociolaboral en la ciudad de Barcelona. Su finalidad fue acercar a los colectivos, asociaciones y personas inmigrantes los distintos servicios de inserción sociolaboral normalizados del Ayuntamiento de Barcelona, a la vez que facilitar la coordinación y adaptación de éstos a las necesidades diferenciales que tiene esta población. El proyecto atendió a un total de 235 personas inmigrantes.

OBJETIVOS DE LA EVALUACIÓN

La finalidad de la evaluación se ha centrado en valorar los efectos del proyecto en los servicios que han participado, en las personas que han sido usuarias directas de las actividades prevista en el proyecto y en las asociaciones y colectivos de inmigrantes que participaron en su desarrollo.

Las cuestiones que interesaron responder con el estudio evaluativo se referían a preguntas tales como ¿qué puede aprenderse del proyecto que sirviera para otros proyectos futuros?, ¿en qué medida el proyecto ha sido útil para satisfacer la necesidad que motivó su creación? ¿era realmente necesario el proyecto?, ¿para quién era necesa-

rio?, ¿cómo deberían funcionar los servicios normalizados para que sean receptivos a las necesidades de los inmigrantes?

En concreto se pretendió:

A. *Análisis de la eficacia del proyecto:* representó una evaluación de resultados centrada en los *objetivos internos* del proyecto. Se analizaron los resultados del proyecto sobre los usuarios de los servicios, las siguientes cuestiones generales orientaron la evaluación:

- ¿En qué medida se han logrado los objetivos previstos para las personas inmigrantes que participaron en él?
- ¿Qué utilidad tuvo el proyecto para estas personas?
- ¿En qué medida las personas que han pasado por el proyecto se han incorporado a la red normalizada de utilización de los servicios de inserción sociolaboral del Ayuntamiento una vez finalizado el proyecto?

B. *Valoración de la efectividad e impacto institucinal del proyecto*: representó una evaluación centrada en los análisis de los efectos del proyecto sobre los servicios de inserción sociolaboral del Ayuntamiento de Barcelona y en las asociaciones de inmigrantes. Se trató de valorar el logro de los objetivos externos del proyecto relacionados con las necesidades sociales que originaron su creación. Las cuestiones de evaluación generales que guiaron este análisis fueron las siguientes:

- ¿Cómo ha incidido el proyecto en el estado de opinión y sensibilidad de las personas responsables los distintos niveles: asociaciones de inmigrantes, personal técnicos y monitores-formadores de los servicios de inserción sociolaboral del Ayuntamiento?
- ¿Cual es el modelo funcional para la inserción sociolaboral de personas inmigrantes que se defiende desde los distintos niveles (usuarios y usuarias individuales, colectivos de inmigrantes y responsables de los servicios de inserción sociolaboral del Ayuntamiento)?
- ¿Qué necesidades surgen y qué modificaciones se plantean en los servicios de inserción sociolaboral del Ayuntamiento para hacerlos receptivos a las necesidades y características peculiares de las minorías de personas inmigrantes?
- ¿Qué necesidades plantean y qué sugerencias ofrecen las asociaciones y personas inmigrantes para acceder a los servicios de inserción sociolaboral?

C. *Realizar una propuesta de* modelo de servicio de inserción sociolaboral que tomara en consideración el aprendizaje y resultados obtenidos en el proyecto Epikouros.

FUNCIÓN DE LA EVALUACIÓN

La demanda explícita que se hace de evaluación obedeció a una función eminentemente sumativa ¿qué se había aprendido de la experiencia una vez finalizada que

pudiera servir para otros proyectos de objetivos similares? Implícitamente la evaluación pretendió cubrir una función socio-política de concienciación del personal de los servicios municipales de las deficiencias institucionales para responder a las exigencias que demandaba la nueva realidad multicultural del mundo laboral. No obstante estas funciones, el planteamiento participativo que se hizo del proceso de evaluación —si bien en la práctica mas limitado que en su diseño original— provocó procesos formativos para algún personal de los servicios de las instituciones que participaron.

PARADIGMA

La evaluación se ha realizado desde un enfoque descriptivo-interpretativo en el cual era básico la reconstrucción de procesos. La parte descriptiva se obtiene a partir de las opiniones de las personas que participaron en el proyecto, bien como usuarias bien como técnicos o responsables de asociaciones, y la interpretativa a partir de las inferencias que se hacen de los significados atribuidos.

MODELO

La evaluación que se presenta puede clasificarse como una evaluación orientada a los objetivos del programa pero no tanto con la finalidad de conocer si se han logrado o no los objetivos del proyecto sino de elaborar una propuesta de atención y asesoramiento en la inserción sociolaboral de inmigrantes a partir de la experiencia obtenida en el proyecto.

METODOLOGÍA

Diseño. Se planteó inicialmente un diseño de evaluación participativo en el que interviniera las principales audiencias: representantes de colectivos de inmigrantes y de personal de las distintos servicios municipales que habían intervenido en proyecto. Consideraciones socio-políticas tanto al inicio como durante el desarrollo limitó considerablemente los niveles de participación previstos.

Técnicas de recogida de información: El plan de recogida de información corresponde a una evaluación de naturaleza descriptiva, donde se toma en consideración los siguientes elementos:

— se identificaron las fuentes de información según, los datos necesarios para responder a las cuestiones de evaluación del estudio: personas inmigrantes usuarias del proyecto, personal de los servicios normalizados de inserción sociolaboral, personal técnico del SOLI (servicio de orientación sociolaboral), los directores y gestores del proyecto y los responsables de los distintos colectivos de inmigrantes que participaron. También fueron fuentes valiosas de información el análisis de los «productos» y documentos generados por el proyecto.

— se seleccionaron las muestras para cada una de las fuentes de información excepto para los responsables y gestores directos del proyecto y responsables de

los colectivos y asociaciones de inmigrantes que fueron tomados en consideración su totalidad. Se utilizaron distintas técnicas de muestreo: aleatorio estratificado para las personas inmigrantes (32 de un total de 200) y opinático por cuotas para los especialistas de los distintos servicios (22 técnicos de los servicios de un total indeterminado).

— estrategias de recogida de información: la entrevista semiestructurada abierta constituyó la fuente fundamental de obtención de información. Se prepararon un total de siete guías de entrevistas atendiendo a las distintas fuentes de información. Se realizaron mediante entrevista personal y en los casos permitidos se llevó a cabo su grabación.

Criterios de valoración: En el Cuadro 1 se recogen los criterios de evaluación que se utilizaron para las finalidades de la evaluación clasificados según fuente de información.

Referencias. Los términos de las comparaciones realizadas fundamentalmente fueron de tipo absoluto, si bien como es de esperar en este tipo de programa, no es que estuvieran establecidos a priori niveles de logro a alcanza. Se tuvieron en cuenta en algunos casos los niveles que se explicitaron en el diseño del programa, los siguientes ámbitos de referencia:

a) Las personas inmigrantes de las minorías étnico-culturales que han participado en el proyecto como usuarios y usuarias de los distintos servicios.
b) El SOLI como servicio específico de orientación para la inserción sociolaboral de inmigrantes promovido por el proyecto.
c) Los diferentes colectivos de inmigrantes que han participado en el proyecto como asociaciones que representan a diferentes minorías étnico-culturales. En concreto fueron: dos colectivos magrebíes (ATIME y Ibn Batuta), el colectivo Dominicano, el Centro Filipino, el colectivo Paquistaní y el colectivo AEJGE de Guinea Ecuatorial.
d) El personal del las actividades y departamentos de los distintos servicios municipales que intervinieron en el proyecto (Servicios Sociales, Barcelona Activa y el Instituto Municipal de Educación de Barcelona)
e) Gestores y diseñadores del proyecto.

Técnicas de registro y análisis de la información. El análisis de la información: se realizó mediante el programa SPSS para el análisis descriptivo univariado y bivariado de los datos cuantitativos, y se usó el programa Etnograph para organizar y codificar los datos en el análisis de contenido de la información cualitativa.

Tipo de decisiones. No se planteó la evaluación asociada a ningún tipo de decisión relativa al proyecto puesto que prácticamente se había acabado cuando se inició la evaluación y no se pretendía repetirlo. Como resultado de la evaluación se realiza la propuesta de un nuevo modelo de servicio que atendiera a las necesidades particulares que tienen las personas inmigrantes en su inserción sociolaboral. La utilidad insti-

CUADRO 1
CRITERIOS DE EVALUACIÓN

ÁMBITOS DE ANÁLISIS	ANÁLISIS DE LA EFICACIA **Resultados respecto objetivos internos**	ANÁLISIS DE LA EFECTIVIDAD E IMPACTO **Resultados respecto objetivos externos**
USUARIOS/AS: Personas inmigrantes que participaron en el proyecto.	• Análisis descriptivo de los usuarios • Participación en los servicios e itinerarios • Valoración del SOLI • Nivel subjetivo de logro • Valoración de servicios utilizados	• Nivel de autonomía para el uso futuro de los servicios del ayuntamiento. • Modelo y funcionalidad de inserción que sociolaboral para inmigrantes.
ASOCIACIONES DE INMIGRANTES	• Grado de conocimiento sobre los servicios de inserción. • Valoración de la experiencia de su participación en el proyecto.	• Cambios y mejoras en servicios para adaptarse a las necesidades de los inmigrantes: modelo funcional de inserción sociolaboral • Valoración y expectativas sobre los servicios.
SERVICIOS SOCIALES: • *SOLI* • *TALLERES PRELABORALES*	• Conocimiento de las asociaciones sobre los servicios de inserción sociolaboral del ayuntamiento • Nivel de logro de los usuarios y usuarias.	• Valoración del modelo de orientación (SOLI) • Efectos del proyecto sobre el uso de los servicios en el futuro por las asociaciones. • Efectos del proyecto sobre los servicios municipales • Coordinación entre servicios • Propuesta de modelo funcional de inserción sociolaboral para inmigrantes • Valoración de la utilidad de sus objetivos • Valoración de su papel en el conjunto de servicios • Modelo de educación multicultural
IMEB *(Clases de castellano)*	• Características de la población atendida y relación con las asociaciones. • Valoración de la experiencia de su participación en el proyecto.	• Cambios/modificaciones a introducir para adaptarse a las necesidades de personas inmigrantes • Efectos del proyecto sobre el servicio
SERVICIOS BARCELONA ACTIVA *(Acciones IOB, Escuelas Talleres, F. Ocupacional. Plan de Ocupación)*	• Grado de conocimiento y valoración del proyecto Epikouros. • Valoración que se hace de las personas inmigrantes y nivel de logro conseguido. • Razones de los resultados positivos y negativos.	• Cambios/modificaciones a introducirse para hacer los servicios más receptivos a las personas inmigrantes • Modelo funcional de inserción sociolaboral para inmigrantes.

tucional de los resultados de la evaluación pueden ser importantes si el modelo de servicio que se propone tuviera los apoyos políticos y gestores necesarios para que fuera tomado en consideración en el desarrollo de una política de inserción sociolaboral de inmigrantes.

Informe de evaluación: Informe escrito y presentación oral de los resultados a los responsables del proyecto. Del informe destaca el capítulo dedicado a la propuesta de un Modelo de servicios de inserción sociolaboral para inmigrante que se extrae de los resultados obtenidos de evaluación (IASI: servicio de información y asesoramiento sobre la inserción sociolaboral). Se intentaron dos presentaciones orales del proyecto atendiendo a los planteamientos de una evaluación participativa: una, con los colectivos de inmigrantes y otras con el personal de las instituciones municipales participantes directamente relacionadas con las actividades del proyecto. Dichas presentaciones constituirían unos espacios privilegiados para el debate sobre el modelo de servicio propuesto, a la vez que reflexión y sensibilización ante los problemas específicos que tiene la población inmigrante para su inserción sociolaboral y las posibles soluciones que pueden abordarse desde las distintas instancias institucionales. Distintas razones limitaron estas presentaciones a una única presentación con los gestores el programa.

6 U 4

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 627-637

UNIVERSIDAD DE CÁDIZ

Olga García Caballer, Mayka García García, Marisol Ibarra Sáiz y Gregorio Rodríguez Gómez
Facultad de Ciencias de Educación. Universidad de Cádiz

TÍTULO DEL PROGRAMA: ESTUDIO COMPARATIVO DE ACCIONES FORMATIVAS PARA JÓVENES MEDIANTE UN PROCESO DE EVALUACIÓN EX-POST, EN EL CONTEXTO DE LAS INICIATIVAS COMUNITARIAS DE EMPLEO

1. INTRODUCCIÓN

El Estudio comparativo de acciones formativas para jóvenes es la concreción de una acción prevista por la Asociación Síndrome de Down de Cádiz y Bahía «Lejeune»[1] integrada en el Proyecto INTERCOM, que se inserta en el marco de la Iniciativa Comunitaria Empleo-Horizon, y contempla diferentes actuaciones dirigidas a la capacitación sociolaboral de jóvenes de la Asociación mediante la experimentación de diversas modalidades de formación profesional: Formación Profesional Ocupacional, Programas de Garantía Social, Segundas Prácticas y Empleo con Apoyo.

Con la inclusión en su proyecto de un estudio comparativo la Asociación se marcó los siguientes objetivos:

- Comparar con suficiente garantía los elementos que se han tenido en cuenta para el diseño de las distintas actividades formativas.
- Evaluar los resultados obtenidos.
- Establecer conclusiones válidas que permitan a las Administraciones responsables adecuar las convocatorias oficiales a aquellos condicionantes que garanti-

1 Para el desarrollo del estudio la Asociación decidió contar con un equipo de evaluación de la Unidad para la Calidad de la Universidad de Cádiz, para lo cual se desarrolló un convenio de colaboración, gestionado por la Fundación Universidad Empresa de la Universidad de Cádiz.

cen una mayor eficiencia en las acciones formativas, y a su vez, resuelvan las necesidades educativas, sociales y laborales de los jóvenes que han desarrollado estas acciones.

El estudio en sí ha sido conceptualizado respondiendo a los interrogantes expresados en la siguiente figura:

<table>
<tr><td colspan="2">• ¿Qué vamos a estudiar? (EL OBJETIVO)</td><td colspan="2">• ¿ Cómo lo vamos a estudiar?</td></tr>
<tr><td colspan="2">Un Programa que consta de actividades formativas para lograr la inserción laboral de jóvenes</td><td colspan="2">Mediante un proceso de EVALUACIÓN</td></tr>
<tr><td colspan="4">Evaluación de programa</td></tr>
<tr><td>¿Quiénes vamos a evaluar?</td><td>¿Para qué vamos a evaluar?</td><td>¿Cómo vamos a evaluar?</td><td>¿Cuándo va a ser evaluado?</td></tr>
<tr><td>Un equipo de Agentes Externos</td><td>Para conocer lo que ha sucedido (productos y procesos) y poner en marcha procesos de mejora</td><td>Sumergiéndonos en el contexto para recoger datos e interpretar conjuntamente lo que está aconteciendo y lo acontecido</td><td>Cuando las acciones han terminado o están terminando</td></tr>
</table>

Figura 1. *Conceptualización de la evaluación en «Estudio Comparativo de Acciones Formativas para Jóvenes».*

Como se desprende de la figura anterior, el «Estudio Comparativo de Acciones Formativas para Jóvenes» (Rodríguez, G.; García, O.; García, M.: 1999) se ha desarrollado como una evaluación externa de programa, de carácter ex - post, desde una óptica cualitativa.

2. JUSTIFICACIÓN DE LA EVALUACIÓN

La tarea de afrontar la evaluación de este proyecto desde la condición de agentes externos al contexto, tomó como punto de partida un proceso de descubrimiento conjunto del objeto de evaluación, cuyo fruto fue el siguiente:

- Acciones formativas en sí mismas (Planificación, Adecuación...)
- Acciones formativas en su desarrollo (Sesiones, Dificultades...)
- Acciones formativas en sus resultados (Resultados, Productos e Impactos)

Este objeto de evaluación y el momento en que la Asociación decide realizarla —cuando la mayor parte de las acciones formativas que constituían el programa habían finalizado o estaban finalizando—, dieron lugar al carácter ex-post-facto de la misma.

En el marco en el que se desenvuelve la Iniciativa Comunitaria Empleo-Horizon, el foco de atención habría que dirigirlo a *Resultados, Productos e Impactos de las Acciones Formativas,* términos que conceptualizamos a partir de lo referido por la Comisión Europea (1997):

— *Productos o realizaciones:* Son los materiales y servicios construidos y directamente producidos por el programa o por el curso. Responden a objetivos operacionales.
— *Resultados*: Es el impacto inicial (o específico) del programa. Responden a la consecución de los objetivos específicos.
— *Efectos*: Es el impacto del programa a largo plazo (o impacto general). Responden a la consecución de los objetivos generales.

No obstante, entendemos que la utilidad de la evaluación —teniendo en cuenta el objeto de la misma— no vendría solamente determinada por la concreción de una serie de datos cuantificables, sino, más bien, por un conocimiento, interpretación y valoración, fruto de un análisis acerca de si los resultados y efectos obtenidos están en relación con los objetivos del proceso diseñado e implementado para la consecución de los mismos, y a su vez por la valía de estos procesos.

La motivación que inspira a la Asociación, para hacer una petición de evaluación, no se reduce a conocer si los métodos de Formación Profesional empleados para la inserción socioprofesional de jóvenes son útiles, sino también a encontrar qué elementos han de ser mejorados, a través de la interpretación acerca de lo que ha acontecido a lo largo de la experiencia. Por todas estas razones se empleó para este estudio una metodología de corte cualitativo, para establecer no sólo los resultados, sino a su vez para fijar y aprender cuales han sido los procesos de formación, ayudando así, a los participantes a comprender sus acciones.

3. JUSTIFICACIÓN METODOLÓGICA

La metodología de evaluación está fundamentada en un proceso de investigación acerca de lo que acontece en este marco, así como en la búsqueda de alternativas de análisis que contribuyan siempre a la calidad de los programas. En nuestro diseño de evaluación destacan tres elementos claves para el análisis metodológico: negociación, contextualización y cualificación.

Negociación	Contextualización	Cualificación
Intercambio de significados e intenciones con las personas que formulan la petición de una evaluación u otro tipo de estudio.	Respuesta a las necesidades específicas de cada uno de los colectivos que demandan una evaluación u otro tipo de estudio.	Facilitar a los implicados en los procesos de evaluación, las herramientas necesarias para el desarrollo propio de un conocimiento e interpretación sobre su contexto.

Estos tres elementos establecen como principio general una humanización de los procesos de evaluación o de cualquier otro estudio relacionado con la calidad de sus acciones. De aquí se genera a su vez un diseño metodológico adecuado a cada uno de los contextos. En nuestro caso desarrollamos, a continuación, el elaborado para la Asociación «Lejeune» en su parte más práctica, centrándonos en las diferentes fases. Para cada una de ellas han sido varias las estrategias utilizadas; trataremos en este apartado de reflejar «el por qué y el cómo» de su elección.

3.1. Diseño Metodológico

El diseño metodológico se estableció en fases con el objeto de establecer la lógica del proceso de evaluación. Éstas fases fueron:

A) *Acceso al campo.*
B) *Descripción de la muestra.*
C) *Temporalización*
D) *Diseño de la recogida de información.*
E) *Análisis de los datos.*
F) *Emisión de Resultados*

A) El ***Acceso al Campo*** es el proceso por el cual nos hemos aproximado a la información relevante para nuestra evaluación; así como el establecimiento de los contactos directos con el personal técnico y directivos de la Asociación, con los cuales se realizaron desde el principio y en todo momento relaciones de información bidireccionales. El Proyecto «Evaluación de la Acciones Formativas para Jóvenes de Lejeune» establecía las bases de acceso a la información, las características generales del tipo de estudio y la evaluación que debíamos realizar. A su vez la documentación base y específica, relacionada con las acciones formativas que la Asociación «Lejeune» poseía ha sido básica para un completo análisis de las acciones desarrolladas.

*B) **Descripción de la muestra***: La selección de diferentes implicados para la muestra realizada por la Asociación, de cara a poder cumplir y desarrollar los objetivos previstos en su proyecto, nos pareció muy acertada, por lo que nos sumamos a ella. Los implicados y su relación para con los proyectos se indican a continuación:

- Alumnos asistentes a las acciones formativas y prácticas de empleo; los destinatarios de las diferentes acciones los constituyen jóvenes con una edad mínima de 16 años, según las características generales exigidas en el programa en la modalidad denominada de integración, lo que supone que a los mismos pueden acceder alumnos con discapacidad y sin ella, así cono jóvenes en situación de deprivación sociocultural. Del total de 61 jóvenes seleccionados, el 50% se incluye en la discapacidad psíquica y el 50% restante son jóvenes en situación de desventaja cultural y económica.

- Monitores de los cursos, encargados de la docencia, asumiendo competencias de formación profesional específicas. Su elección fue cubierta sobre la base de la experiencia en las distintas acciones. El número de monitores fue de 13, y su distribución fue de 7 en formación ocupacional y 6 en formación básica.
- Mediadores sociolaborales, presentes en todo el proceso de integración de la persona con necesidades especiales; una de sus principales funciones es la de dinamizar en los procesos de inserción de los jóvenes. La formación de estos mediadores se realizó a través de un curso impartido por la misma Asociación «Lejeune». Participaron un total de 10 mediadores.
- Familias de los alumnos, uno de los pilares fundamentales para la total integración de estos jóvenes y el máximo aprovechamiento de las diferentes enseñanzas. Estuvieron implicados un total de 61 familias.
- Tutores laborales de las prácticas en la formación y de las prácticas laborales, personal de las empresas o Administraciones que constituyen el referente de los jóvenes en sus prácticas. Desarrollaron funciones de tutorización 34 personas.
- Empresarios y Administraciones Públicas donde los jóvenes realizan sus prácticas de formación y que también son beneficiarios indirectos de los procesos de formación en los jóvenes. En el proceso había 11 empresarios involucrados.

Resulta significativo resaltar que, dadas las características numéricas de la población implicada en el proceso de evaluación, la muestra resultante alcanzó la totalidad de la citada población.

C) La ***Temporalización*** de los procesos permite a las partes, conocer el momento preciso en el que las acciones de evaluación se encuentran, y sobre todo adquiere un compromiso de acción abierta al continuo proceso de negociación entre las partes. En este estudio las actividades establecidas en la temporalización según las características del trabajo a realizar fueron:

Mayo	*Acuerdo con la Asociación en diseño de evaluación*
Junio	*Preparación de instrumentos y estrategia de recogida de datos*
Junio/Julio	*Recogida de datos*
Julio/Agosto	*Análisis de datos*
2ª sem. Agosto	*Elaboración de informe preliminar*
1ª sem. Sept.	*Discusión de informe preliminar*
3ª sem. Sept.	*Elaboración de informe definitivo*
4ª sem. Sept.	*Presentación de informe definitivo*

D) ***Diseño de la recogida de información***: Previa a la recogida de información y la consiguiente construcción de instrumentos a tal efecto, se inició un proceso de con-

creción de dimensiones y categorías, dadas las características peculiares del contexto. Otro elemento decisivo para llevar a cabo la citada construcción es la existencia escasa de indicadores —los existentes además son de corte cuantitativo— para la evaluación de estos programas en el marco de las Iniciativas Comunitarias de Empleo.

Una primera revisión documental, nos condujo a la delimitación de cuatro momentos relevantes en el transcurso de las acciones formativas, de las prácticas laborales y de los procesos de inserción sociolaboral. Estos momentos fueron tomados como dimensiones generales:

- *Contexto y justificación* como momento preparatorio y de diseño.
- *Inicio* como momento de planificación una vez puestas en marcha las acciones.
- *Desarrollo* como procesos de implementación de las acciones.
- *Resultados, efectos e impacto* como productos resultantes de las acciones desarrolladas a lo largo del Proyecto INTERFACE, INTERCOM y los Programas de Garantía Social.

A continuación se desglosó cada una de las dimensiones en categorías relevantes para la recogida de información y el análisis de la misma. Dado que el proceso de evaluación era a su vez un estudio que debía permitir la comparación entre las diferentes fórmulas de formación para la inserción laboral de jóvenes; resultó un modelo categorial que permitía por un lado descubrir los elementos característicos de cada una de ellas, a la vez que aportaba elementos clave para la comparación.

Sistema de Categorías General

1. ASDOWN
2. MOTIVACIÓN PARA LA EVALUACIÓN
3. CONTEXTO Y JUSTIFICACIÓN DE INTERFACE
4. CONTEXTO Y JUSTIFICACIÓN DE PGS
5. CONTEXTO Y JUSTIFICACIÓN DE INTERCOM
6. CONTEXTO Y JUSTIFICACIÓN DE LAS ACCIONES FORMATIVAS
7. PLANIFICACIÓN DE LA ENSEÑANZA:
8. DESARROLLO DE LA ENSEÑANZA
9. EVALUACIÓN DE LA ENSEÑANZA
10. RELACIONES
11. PRÁCTICAS EN EMPRESA
12. EMPLEO CON APOYO
13. PROPUESTAS DE MEJORA

Partiendo de este sistema de categorías y atendiendo a las peculiaridades del contexto, de la muestra, y del programa, se diseñaron las herramientas e instrumentos evaluativos que a continuación se describen.

610

- *Análisis de Documentos*. Para el análisis de estos documentos se realizó una primera lectura, una toma de contacto con las diferentes programaciones, memorias, proyectos, artículos de prensa, etc. La cantidad de información y sobre todo la ordenación inicial en los documentos aportados por la Asociación, nos llevo a realizar un diseño de carpetas con todo el material de las acciones formativas para poder realizar un posterior análisis más exhaustivo de los datos a analizar. Posteriormente el proceso de recogida de datos de los documentos se realizó en base al primer sistema de categorías. A través de un diseño de plantillas de entrada de categorías, se iba apuntando un código numérico adjudicado previamente a cada documento, el cual nos indicaba dónde se hallaba la información más relevante en nuestro estudio para cada una de las acciones formativas.
- *Cuestionarios*: Dada la diversidad de información que queríamos recabar, las características especificas de cada una de los informantes con los que queríamos contactar, y sobre todo para conseguir en los resultados un mayor margen de fiabilidad, se realizaron dos tipos de cuestionario: *Cuestionario Telefónico* y *Cuestionario Presencial*.

Con relación a los Cuestionarios Telefónicos se procedió a su realización después del análisis de documentos, y de los contactos mantenidos con la Asociación «Lejeune» y los agentes implicados.

El *Cuestionario Telefónico*, dividido a su vez en dos, estaba destinado a los tutores laborales y a los empresarios de los jóvenes; se estructura en torno a los siguientes criterios de: Contexto, Desarrollo de las Prácticas, Evaluación, Relaciones, inserción laboral y Propuestas de Mejora.

El *Cuestionario Presencial* destinado a las familias de los jóvenes participantes de las acciones formativas se estructura sobre la base de: Vinculaciones, Valoraciones de las Acciones Formativas, Docencia y Seguimiento, Valoración de las Prácticas Laborales, Valoración del Empleo con Apoyo y Evaluación.

- *Entrevistas Semiestructuradas*. Se diseñaron dos tipos de *entrevistas* semiestructuradas, una de *carácter individual*, destinada a los alumnos y otra de *carácter grupal* destinada a los monitores y mediadores. La peculiaridad de la *entrevista grupal* radicaba en establecer la recogida de información con las todas los docentes al mismo tiempo; creemos que así las valoraciones a realizar serán mucho más ricas dado que su trabajo no se desarrolló por separado. Los aspectos tratados para la entrevista grupal fueron: Actividad Profesional, Alumnos, Planificación de la enseñanza, Desarrollo del curso, Evaluación, Relaciones, Prácticas Formativas y Laborales, Empleo con Apoyo y Necesidades.

Con relación a la *Entrevista Semiestructurada individual* al alumno, se centró en: Acceso al curso, Desarrollo de los cursos, Profesorado, Evaluación, Relaciones y Necesidades (Valoraciones).

*E) **Análisis de los datos***: Para el análisis de la información se siguió el proceso planteado por Rodríguez, Gil y García (1996), a través de tres etapas: reducción de datos, obtención de resultados y verificación de conclusiones. El análisis de los datos se realizó utilizando ficheros de transcripciones de las entrevistas y los resultados del cuestionario —para que posteriormente fueran reconocidos por el programa NUDIST 4.0 (Rodríguez y otros, 1995)—. En la primera etapa de reducción de datos se identificaron y clasificaron las unidades relevantes y significativas a través de la categorización y codificación, para lo cual se adoptó el sistema general de categorías que tomo como base la Tabla de Categorías diseñada para el análisis de documentos, y codificación del material de las entrevistas y cuestionarios. De cara a facilitar el proceso de categorización para su análisis, se diseñó un código de letras y números con el objeto de clasificar la información y establecer unos códigos que posteriormente aparecerán en el informe.

En una segunda etapa se procedió a la transformación de los datos seleccionando la información que se había clasificado bajo las definitivas categorías de análisis, y el resultado de la misma fue objeto del análisis por parte del equipo de evaluación, en el intento de determinar la valoración y opinión —obtención de resultados— de cada uno de los diferentes agentes de las acciones formativas.

*F) **Emisión de Resultados***. Por último se redactó el informe del desarrollo de cada una de las acciones formativas y que constituye en sí mismo la verificación de las conclusiones que pudieran establecerse.

4. RESULTADOS

En este artículo se ha descrito un proceso de evaluación externa y un estudio comparativo de acciones formativas, relatando la experiencia a través de sus bases metodológicas. Se ha desarrollado una metodología *sistemática*, según un plan —en parte— previamente fijado y *flexible*. Hemos querido establecer estas dos premisas en todo momento en el estudio, pero tal vez sea en el apartado metodológico donde con más sentido se señalan. Esta flexibilidad metodológica también se ha visto reflejada en la valoración y comparación de los resultados. Desde la óptica cualitativa bajo la cual se ha desarrollado la evaluación, los resultados de este estudio para la evaluación ex-post se concretaron en:

— Puntos fuertes y débiles del modelo formativo de la Asociación «Lejeune», y recomendaciones para la mejora.
— Descripción, Interpretación y Valoración Global de las acciones Formativas. (Análisis Comparativo)

Para estos dos elementos, el nexo unificador fue el sistema de categorías empleado en toda la evaluación, el cual fue aportando, a través de la recogida de la información, las conclusiones relevantes a nuestro estudio. La dificultad de este caso es que nuestra evaluación es al mismo tiempo un estudio comparativo y una evaluación ex-post, por

lo que las conclusiones se identificarán de una manera o de otra. Para el primer elemento se ofrece siguiendo estos cuatro aspectos:

- *El modelo general*
- *Las acciones formativas.*
- *Segundas prácticas.*
- *Empleo con Apoyo.*

En cambio, para el segundo elemento de valoración global de las acciones formativas, lo que propiamente podemos llamar los resultados del estudio comparativo han buscado indicadores propios de comparación:

— *Integración.*
— *Modelo Formativo.*
— *Diseño Didáctico.*
— *Alumnos.*
— *Familias.*
— *Empresas.*
— *Figuras profesionales.*
— *Nivel Formativo.*
— *Nivel Personal.*
— *Nivel De Integración.*
— *Contrato Laboral.*

5. CONCLUSIONES

Para finalizar las conclusiones metodológicas que este equipo de evaluación destaca con respecto a este trabajo:

- Los diferentes pasos seguidos a lo largo de este proceso nos ha permitido sumergirnos en el contexto singular de la Asociación «Lejeune», proporcionando valiosas claves para la interpretación de lo que en el mismo aconteció. Nuestra integración en el contexto facilitó el desarrollo de un clima de trabajo caracterizado por la transparencia y el intercambio mutuo de información.
- Así mismo, este equipo de evaluación a lo largo del transcurso de su trabajo con la Asociación «Lejeune», ha detectado un cambio en la concepción que estos asumían, —ideas previas de los procesos evaluativos—; empezando a pensar por sí mismos en autoevaluación —reconstrucción de pensamientos—.
- Al desarrollo del aspecto señalado con anterioridad ha contribuido en gran medida la utilización de la entrevista semiestructurada como herramienta evaluativa, que ha aportado, no sólo información rica y relevante sino que además refuta la importancia de la fase de diseño de evaluación y la elección adecuada

de las técnicas e instrumentos. Además la entrevista ha posibilitado la humanización de evaluación en el contacto cara a cara, descubriendo y compartiendo significados.

- En este tipo de metodología de evaluación, el evaluador debe adquirir un fuerte compromiso de trabajo, que entre otras cosas supone: «hablar el mismo lenguaje» utilizado en el contexto, y dominar las claves como elemento esencial para el desarrollo de la comunicación.
- La insuficiencia de indicadores de evaluación en el marco de las Iniciativas Comunitarias de Empleo nos lleva a la necesidad de construirlos, o como en nuestro caso, construir otros «criterios» en los que centrar la evaluación.
- En relación con lo anterior, el carácter estrictamente cuantitativo de los mismos (nº de beneficiarios, nº de inserciones...), así como su foco de atención en productos, nos conducen necesariamente a investigar en otros modelos de evaluación, como el propuesto a través de evaluación respondente.
- Llevar a cabo un proceso de evaluación de procesos y productos desde una óptica cualitativa supera el modelo evaluativo practicado en el marco de referencia, ya que el uso de la evaluación deja de ser simplemente sumativo, para transformarse en formativo de los contextos.

A lo largo de esta experiencia de evaluación hemos aprendido que esta opción no supone simplemente un compromiso de sus profesionales; para que de verdad se impulsen procesos de mejora, los evaluados deben sentirse protagonistas y partícipes a lo largo del proceso de evaluación. Creemos que la clave reside en empezar a abandonar nuestra posición de «expertos» para tomar conciencia de la necesidad de ser facilitadores de herramientas para comprender la realidad.

6. BIBLIOGRAFÍA

Arnal, J., Del Rincón, D. y Latorre A. (1994). *Investigación Educativa. Fundamentos y Metodologías.* Barcelona: Labor.

Comisión Europea (1997). *Evaluating EU Expediture Programmes: A Guide. Ex post and intermediate Evaluation.* Luxemburgo: Oficina de Publicaciones de las Comunidades Europeas.

Fitz-Gibbon, C. (1987). *Evaluator´s Handbook.* Londres: SAGE.

Joint Committee for Standards for Educational Evaluation (1998). *Normas para la evaluación de programa, proyectos y material educativo.* México: Trillas.

House, E.R. (1994). *Evaluación, ética y poder.* Madrid: Morata.

OCDE (1991). *La Evaluación de Programas de Empleo y de medidas sociales.* Madrid: Centro de Publicaciones Ministerio de trabajo y Seguridad Social.

Patton, M. (1996). *Utilization- Focused Evaluation.* Londres: Sage.

Rodríguez, G. y Otros (1995). *Análisis de datos cualitativos asistidos por ordenador: AQUAD y NUDIST.* Barcelona: PPU.

Rodríguez, G., Gil, J. y García, E. (1996). *Metodología de la Investigación Cualitativa.* Archidona: Aljibe.

614

Simons, H. (1999). *Evaluación democrática de instituciones escolares.* Madrid: Morata.
Stake, R. (1975). *Evaluating the Arts in Education: A responsive Approach.* Columbus, OH: Charles E. Merrill.
Stake , R. (1995). *The Art of Case Research.* Newbury Park: Sage.
Stufflebeam, D. y Shinkfield, A. (1987): *Evaluación Sistemática: Guía teórica y práctica.* Barcelona: Paidós.

615

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 639-645

UNIVERSIDAD COMPLUTENSE

Mercedes García García
Departamento de Métodos de Investigación y Diagnóstico en Educación

TÍTULO DEL PROGRAMA: METODOLOGÍAS UTILIZADAS EN LAS EVALUACIONES DE PROGRAMAS DEL DEPARTAMENTO DE MÉTODOS DE INVESTIGACIÓN Y DIAGNÓSTICO EN EDUCACIÓN DE LA UNIVERSIDAD COMPLUTENSE DE MADRID

La evaluación de programas educativos se inicia como línea de investigación en el Departamento de M.I.D.E. de la Universidad Complutense a mediados de los años ochenta, dando lugar a gran parte de los cursos del programa de doctorado *Diagnóstico, Medida y Evaluación de la Intervención Educativa* de la década posterior. En la síntesis que a continuación presentamos se recogen diez evaluaciones en las que han participado profesores del departamento, siete de ellas finalizadas y tres en proceso de realización.

TEMÁTICA EVALUADA. Tomando el concepto de programa en su versión más amplia, se han evaluado programas y acciones formativas tanto en el ámbito educativo como en el sanitario, abarcando las etapas de educación infantil, educación primaria, educación secundaria, educación universitaria, formación ocupacional y formación continua.

En la mayoría de los trabajos se ha actuado como evaluadores externos al programa, excepto en cuatro proyectos en los que es el propio equipo evaluador quien a la vez diseña o es responsable del programa a evaluar.

Los títulos, responsables y objeto evaluado se resumen en la Tabla 1 (el orden de presentación utilizado es el del año de realización de la evaluación):

TABLA 1

	Título del programa	Responsable del programa	Responsable de evaluación	Objetivo evaluado
1	*Programa de Formación Ocupacional* 1984-1989	Instituto de Molinería e Industrias Cerealísticas	Rafael Carballo	Formación Ocupacional
2	*Estrategias adaptativas para alumnos de bajo rendimiento* 1995-96	IES Palomeras-Vallecas MIDE UCM	María Castro Mercedes García Emelina López Inmaculada Asensio	Adaptación curricular (Biol., Física y Lengua) (3º y 4º ESO)
3	*Educación Infantil* 1995-1997	Ministerio de Educación y Cultura	Mercedes García INCE	Educación Infantil (LOGSE)
4	*Programa de intervención en Educación para la Salud* 1996-1998	Narciso García Nieto Mª José Fernández Rafael Carballo	Narciso García Nieto Mª José Fernández Rafael Carballo	Programa de educación para la salud (3º y 4º ESO)
5	*Necesidades formativas en la Industria Harinera y de Sémolas* 1997	Instituto de Molinería e Industrias Cerealísticas FORCEM	Ramón Sánchez (coor) Ángel Jiménez L. Roces Rafael Carballo R.Cortés	Formación continua: análisis de necesidades
6	*Apoyo y refuerzo de habilidades y aprendizajes instrumentales* 1999-2000	Narciso García Nieto Rafael Carballo Mª José Fernández	Rafael Carballo Mª José Fernández Narciso García	Programa de apoyo escolar (Ciclo Inicial, Ed.Primaria)
7	*Prevención y detección precoz del VIH-Sida* 1999-2000	Pedro Montilla (Consejería de Salud, CAM)	Rafael Carballo Chantal Biencinto	Acciones formativas (Servicios de atención primaria, CAM)
8	*PIPEI: Programa de Intervención Preventiva en Educación Infantil* 2000-2001	Paula Zamorano (Colegio Sagrado Corazón. Fuencarral)	Laura Oliveros Covadonga Ruiz	Programa PIPEI (Ed. infantil)
9	*Rentabilidad de la Formación Continua en el empleo* 2000-2001	UGT-UCM	Rafael Carballo	Formación continua (trabajadores Sector servicios)
10	*Calidad de la Enseñanza Universitaria ofrecida*	J.M. García Ramos Dir. Centro Universitario Francisco de Vitoria Clemente López	J.M. García Ramos Elvira Congosto Jesús Alcalá	Proyecto educativo Servicios Docencia Investigación

617

OBJETIVOS Y FUNCIONES DE LA EVALUACIÓN

Los estudios que presentamos se podrían encuadrar en tres grupos en función del objetivo prioritario que ha dirigido la evaluación. En el grupo más numeroso, el objetivo principal ha sido contrastar si las metas del programa se alcanzaron, analizando los cambios, mejoras e impactos deseables en los beneficiarios directos o indirectos de la intervención. En un segundo grupo ha sido el análisis de necesidades el que ha guiado la evaluación, seguido frecuentemente con la formulación de acciones educativas. Por último, el objetivo de dos estudios ha sido la evaluabilidad del objeto. De forma más detallada se presentan en la Tabla 2.

Las funciones sumativa, centrada en la evaluación de los resultados, y formativa, centrada en el seguimiento y mejora del propio programa, son las que se citan con más frecuencia, excepto en dos casos en los que es la evaluación de necesidades y la información sociopolítica la que determina la evaluación.

PARADIGMA Y MODELO DE LA EVALUACIÓN

La mayor parte de las evaluaciones se han inscrito en el paradigma cuantitativo, aunque la mitad de estudios añaden el cualitativo como aspecto complementario. Así mismo, el modelo dominante es el de resultados si bien se observan proyectos con elementos procesuales y de entrada, ya sea referidos al contexto, a las necesidades o al propio diseño del programa a evaluar (Tabla 3).

METODOLOGÍA DE LA EVALUACIÓN

El diseño que se ha utilizado con más frecuencia ha sido el cuasi-experimental (con un grupo de intervención y medidas tomadas antes y despues del tratamiento; en pocos casos se ha utilizado grupo de control o de comparación) y en su defecto, el diseño preexperimental de carácter descriptivo o correlacional, con medidas recogidas durante y/o tras la intervención. El cuestionario es la técnica de recogida de información más empleada, seguida de la entrevista y de la observación.

Los criterios utilizados para valorar la información son, principalmente, la eficacia y la eficiencia y, puntualmente, la viabilidad, la funcionalidad y la calidad de procesos. No todas las evaluaciones plantean referencias, las que lo hacen suelen emplear criterios prefijados, pero lo más frecuente es valorar la mejora, el cambio o el progreso sobre el propio programa.

Todos los estudios presentan análisis descriptivos y muchos de ellos emplean el análisis de varianza como técnica de contraste. Por último, las decisiones suelen hacer referencia, por un lado, a la mejora, modificación o continuidad del programa o, por otro lado, a la descripción de la situación evaluada o a la viabilidad del modelo o proceso de evaluación (Tabla 4).

TABLA 2

	Título	Objetivos de la evaluación	Función
1	*Evaluación de un programa de Formación Ocupacional* 1984-1989	• Determinar la eficacia de los cursos impartidos • Detección de problemas en el desarrollo de los cursos	Sumativa Formativa
2	*Estudio del bajo rendimiento y de las respuestas educativas para el alumno que no alcanza los objetivos mínimos...* 1995-96	• Analizar la viabilidad de las estrategias de adaptación diseñadas en contexto regular • Determinar la eficacia de las estrategias a través de la mejora de los resultados, actitudes y la satisfacción de los implicados	Formativa Sumativa
3	*Estudio de la viabilidad de la evaluación de la Educación Infantil* 1995-1997	• Analizar la viabilidad del modelo INCE en la evaluación de la educación infantil • Conocer ajuste directrices reforma y expectativas de los mediadores educativos • Valorar dimensiones del entorno, proceso y resultados asociados a la etapa	Sociopolítica
4	*Eficacia de un programa de intervención en Educación para la Salud* 1996-1998	• Detectar las necesidades de formación de los alumnos de secundaria en educación para la salud • Analizar el funcionamiento del programa	Formativa
5	*Estudio de nuevas necesidades formativas en la Industria Harinera y de Sémolas* 1997	• Detección de nuevas tendencias productivas • Elaboración de cursos de formación ajustados a las necesidades productivas	Análisis necesidades
6	*Apoyo y refuerzo de habilidades y aprendizajes instrumentales en niños con dificultades escolares* 1999-2000	• Determinar la eficacia del programa a través de la mejora de los resultados en tests estandarizados	Formativa
7	*Prevención y detección precoz del VIH-Sida en los Servicios de Atención Primaria de la Comunidad de Madrid* 1999-2000	• Determinar el impacto de los cursos en la detección precoz y prevención del VIH-Sida en la población que accede a los servicios de Atención Primaria de la CAM	Formativa
8	*PIPEI: Programa de Intervención Preventiva en Educación Infantil* 2000-2001	• Favorecer la adaptación de los niños en el momento de la escolarización • Mejorar el proceso de enseñanza-aprendizaje en el primer año de escolarización • Crear una relación de colaboración familia-escuela • Valorar el ajuste entre recursos empleados y beneficios obtenidos • Valorar la coherencia interna entre los distintos elementos del programa	Sumativa Formativa
9	*Estudio de evaluación de la rentabilidad de la Formación Continua en el empleo de los trabajadores del Sector Servicios* 2000-2001	• Determinar la eficacia y rentabilidad laboral de la Formación Continua • Elaborar una guía de Buenas Prácticas para la mejora de la Formación Continua en la Negociación Colectiva	Sumativa Formativa
10	*Evaluación de la calidad de la Enseñanza Universitaria ofrecida*	• Mejora institucional	Formativa

TABLA 3
PARADIGMAS Y MODELOS DE EVALUACIÓN

	Título	Enfoque	Modelo
1	*Evaluación de un programa de Formación Ocupacional*	Cuantitativo + Cualitativo	Proceso-Producto
2	*Estudio del bajo rendimiento y de las respuestas educativas para el alumno que no alcanza los objetivos mínimos...*	Cuantitavo + Cualitativo	Contexto + Entrada + Proceso + Resultados
3	*Estudio de la viabilidad de la evaluación de la Educación Infantil*	Cualitativo (análisis de documentos) + Cuantitativo	Contexto (político + curricular) + Entrada + Proceso + Resultados
4	*Eficacia de un programa de intervención en Educación para la Salud*	Cuantitativo	Resultados
5	*Estudio de nuevas necesidades formativas en la Industria Harinera y de Sémolas*	Cuantitativo + Cualitativo	Entrada (necesidades)
6	*Apoyo y refuerzo habilidades y aprendizajes instrumentales en niños con dificultades...*	Cuantitativo	Resultados
7	*Prevención y detección precoz del VIH-Sida en los Servicios de Atención Primaria...*	Cuantitativo	Resultados
8	*PIPEI: Programa de Intervención Preventiva en Educación Infantil*	Cuantitativo	Entrada (necesidades + diseño) + Proceso + Producto
9	*Estudio de evaluación de la rentabilidad de la Formación Continua...*	Cuantitativo + Cualitativo	Resultados
10	*Evaluación de la calidad de la Enseñanza Universitaria...*	Cuantitativo + Cualitativo	CIPP adaptado

620

TABLA 4

	Diseño	Instrumento	Criterios	Referencia	Análisis	Decisión
1	Cuasi-experim Pre-Post	Cuestionario Entrevista P.Rendimiento	Eficacia Eficiencia		Descriptivos ANVA A.Correspond	Modificaciones en estructura programa
2	Cuasi-experim Pre-Post + control	Cuestionario P.rendimiento Entrevista	Eficacia Viabilidad	Mejora post y superioridad CriteriosTeoría	Descriptivos ANCOVA	Introducir reajustes y aplicar siguiente curso
3	Descriptivo	Cuestionario A.Documentos	Viabilidad	Nacional (LOGSE)	Descriptivos A.Factorial A.Correlación	Reajustes modelo, instrumentos y procedimento
4	Cuasi-experim Pre-Post	P.objetiva Cuestionario	Eficacia Efectividad	Mejora post Criterial	Descriptivos ANVA ANCOVA Prueba t	Intervenir antes de 3º ESO Necesidad de intervención Dificultad cambio actitudes
5	Descriptivo	Cuestionario Entrevista			Descriptivos	Elaboración cursos de formación
6	Cuasi-experim Pre-postest + grupo compar	Test ABC-MAE	Eficacia Efectividad	Mejora post	Descriptivos ANVA ANCOVA	Usar material de apoyo en cursos sucesivos
7	Cuasi-experim	Cuestionario Observación	Viabilidad Eficacia	Autoreferencia	Descriptivos Multivariada	No viable impacto medio Cambios en el procedimiento
8	Cuasi-experim	Cuestionario Entrevista Observación	Eficacia Eficiencia Funcionalidad	Criterial	Descriptivos Correlaciones	Continuidad, modificación o supresión del programa
9	Cuasi-experim	Cuestionario Entrevista Grupos de discusión	Eficacia Eficiencia		Descriptivos ANVA	Elaboración de guía
10	Validación instrumentos	Descriptivo + Correlacional	Cuestionario Entrevista Observación	Criterios calidad (P+R)	Eficacia	Efectividad Criterial (libre o prefijado)

INFORME DE EVALUACIÓN

La extensión de los informes es variada (entre 6 y 160 páginas), aunque lo más frecuente es que esté entre 40 y 50 páginas. Los destinatarios del informe se reparten a partes iguales entre entidades públicas (Instituto Nacional de Calidad y Evaluación, Sindicatos, Ministerio de Agricultura, Ministerio de Educación, Universidad Complutense, Consejería de Salud de la C.A.M., Oficina Regional de Acciones frente al SIDA) y privadas, siendo habitual en este caso que se dirijan a los responsables e implicados de los centros y servicios donde se desarrolló el programa (equipos directivos —académicos o de servicios administrativos—, profesores y padres) o a empresarios del sector evaluado y se centre en la presentación de resultados y, en algunos casos, se prescriban las acciones a emprender. En una y otra situación suelen tener una difusión limitada.

BIBLIOGRAFÍA

Carballo, R. (1989). Evaluación de un programa de Formación Ocupacional. *Tesis doctoral, Madrid: Universidad Complutense.*

Carballo, R., García Nieto, N. y Fernández, M.J. (2000). *Apoyo y refuerzo de habilidades y aprendizajes instrumentales en niños con dificultades escolares.* Madrid: UCM.

Carballo, R. y Biencinto, Ch. (2000). *Prevención y detección precoz del VIH-Sida en los Servicios de Atención Primaria de la Comunidad de Madrid.* Madrid: Oficina Regional de Acciones frente al SIDA, Consejería de Salud.

Carballo, R. (en proceso). *Estudio de evaluación de la rentabilidad de la Formación Continua en el empleo de los trabajadores del Sector Servicios.* Dpto. MIDE – UGT-UCM.

Castro, M., García, M., López, E., Asensio, I., Monge, C., Martínez, J. Calero, J.M. y Juiz, M.T. (1996). *Bajo rendimiento y atención a la diversidad en Educación Secundaria Obligatoria: estudio de los factores determinantes del bajo rendimiento y de las respuestas educativas para el alunado que no alcanza los objetivos mínimos y propuestas de mejora.* Madrid: MEC. Informe no publicado.

García, M. (1997). *Estudio de la viabilidad de la evaluación de la Educación Infantil.* Madrid: INCE. Informe no publicado.

García Ramos, J.M. (2000). *Evaluación de la Calidad de la Enseñanza Universitaria del Centro Universitario Francisco de Vitoria.* Madrid: C.U. Francisco de Vitoria.

García Nieto, N., Fernández, M.J. y Carballo, R. (1998). *Eficacia de un programa de intervención en Educación para la Salud.* Dpto. MIDE, UCM.

Oliveros, L. y Ruiz, C. (en proceso). *PIPEI: Programa de Intervención Preventiva en Educación Infantil.* Dpto. MIDE - Colegio Sagrado Corazón de Fuencarral.

Sánchez, R. (coord.) (1997). *Estudio de nuevas necesidades formativas en la Industria Harinera y de Sémolas.* Madrid: Instituto de Molinería e Industrias Cerealísticas.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 647-651

UNIVERSIDAD DE GRANADA

Leonor Buendía Eisman, Eugenio Hidalgo Diez, Beatriz García Lupión, Daniel González González, Rafel López Fuentes, Eva María Olmedo Moreno.

TÍTULO DEL PROGRAMA: **LA MEJORA DE LA CALIDAD EN CENTROS EDUCATIVOS**

RESPONSABLE DEL PROGRAMA:

Entidad: C. «SAFA» (SAGRADA FAMILIA- ALCALÁ LA REAL) JAEN.
Director del Centro: Lorenzo Rus
Jefe de Estudios E. Secundaria: Mª Antonia Sánchez
Jefe de Estudios Bachillerato y Ciclos: Javier Cano

OBJETO EVALUADO:

Dinamización, proyecto de centro, recursos, procesos, satisfacción con el centro, satisfacción del profesorado, impacto en la sociedad y resultados del centro.

OBJETIVO DE LA EVALUACIÓN

Conseguir una cultura de la evaluación en el centro encaminada a la mejora global y a la satisfacción personal del trabajo realizado.

Este objetivo general lo conseguiremos a través de los siguientes objetivos parciales:

a) Generar hábitos de reflexión sobre las prácticas docentes, orientadoras y de gestión de los diferentes estamentos del centro.

b) Confección de instrumentos, surgidos desde las necesidades de nuestro centro, para poder recoger información que nos ayude a la toma de decisiones.
c) Conseguir un mejor nivel formativo de los profesores del centro, a través de la implicación y colaboración en el proceso.
d) Establecer colaboraciones entre los diferentes niveles educativos (no universitarios-universitarios) que permita el enriquecimiento mutuo y el asesoramiento de evaluadores externos.
e) Vistos los puntos débiles de la institución, establecer los aspectos de mejora, para que, tras esta cultura de la evaluación, se puedan conseguir estandares de calidad, a los que debe tender el centro.

Función: Formativa.

Paradigma: Cualitativo.

Modelo: Proceso-producto.

METODOLOGÍA DE LA EVALUACIÓN

En la metodología a seguir en el desarrollo del programa se establecen las siguientes etapas:

1. Autorreflexión del profesorado sobre la necesidad de evaluar el centro como elemento de mejora: etapa diagnóstica.
2. Autoevaluación.
   - Primera fase: Organización del trabajo:
     — Creación del equipo de autoevaluación con miembros integrantes del centro.
     — Formación de los componentes del equipo en una cultura de la autoevaluación y de la gestión de la calidad.
   - Segunda fase: Trabajo de campo:
     — Recogida de información.
     — Establecer grupos de reflexión.
     — Redacción del autoinforme: puntos fuertes y puntos débiles.
     — Sugerencias de mejora.
3. Evaluación externa.
   - Lectura y estudio del informe de autoevaluación.
   - Entrevistas a todos los implicados en el proceso: equipo directivo, profesores/as, alumnos/as, padres/madres y personal laboral.
   - Audiencia pública, donde participarán todos los sectores que configuran el centro.
   - Redacción de un informe a la vista del autoinforme y de las visitas realizadas, donde se hará constar: puntos fuertes y puntos débiles del funcionamiento del centro. Listado y explicación de sugerencias de mejora.

4. Remitir el informe al equipo de autoevaluación.
5. Informe final: será realizado por el equipo de evaluación externa y los profesores del centro.
6. Formación de un equipo de seguimiento, para dinamizar y continuar en sesiones de autorreflexión-evaluación.

Técnicas de recogida de información

Las técnicas que utilizaremos son:

a) Grupos de discusión para las fases de reflexión.
b) Entrevistas y cuestionarios para la fase de diagnóstico y autoevaluación que serán realizados por el profesorado, familias y alumnado, con el asesoramiento de los profesores universitarios del Área de Métodos de Investigación y Diagnóstico en Educación.
c) Análisis de Contenidos de los documentos del centro (diarios, proyectos de centro) y análisis del discurso, en las sesiones conjuntas del profesorado interniveles.

Criterios utilizados para valorar la información

Los criterios utilizados han sido:

Suficiencia. Tras la reflexión de los profesores/as la información recibida la considerabamos insuficiente. En función de los estándares establecidos, procedimos a realizar entrevistas y/o cuestionarios.

Viabilidad y efectividad. Este criterio se ha conseguido a través de reuniones previas, donde los profesores discutían la viabilidad o no de las cuestiones planteadas por nosotros en el centro, a la vez que valoraban la efectividad de las dimensiones diseñadas como elemento de mejora.

Evaluabilidad. Igualmente hemos tenido en cuenta el criterio de evaluabilidad, a través de un equipo de seguimiento, para dinamizar y continuar en sesiones de autorreflexión, una vez puestos en marcha los planes de mejora.

Referencias

Como punto de partida analizamos el modelo EFQM, encontrando grandes inconvenientes para su adaptación al contexto escolar y dificultad de interpretación del vocabulario utilizado.

De esta forma hemos realizado un recorrido por la literatura sobre el tema para, a partir de ella, proponer un modelo propio que nos permita contemplar los aspectos contextuales y las realidades educativas de nuestro centro.

El primer problema es la propia definición de calidad. Los valores de quien la define, marca el desarrollo del concepto.

Comenzando por el principio del interés por evaluar la calidad de un centro educativo, nos encontramos con los establecidos por el *Libro Blanco de 1969*.

Igualmente y abundando en los mismos aspectos, apareció *el II Informe Foessa de 1970*. A este le sucedió el denominado *Informe de evaluación de 1976* y a éste le siguieron otros como: *el III Informe Foessa de 1978, el documento de 1981* sobre enseñanza secundaria; el *IV Informe Foessa*, publicado dos años más tarde, que recoge las mismas anomalías que el informe de 1981.

El V Informe Foessa fue publicado en 1994, y recoge críticas que, en parte, son coincidentes, con las que se recogen en el libro blanco de 1989 y en el documento sobre *Centros educativos y calidad de la enseñanza: propuesta de actuación* publicado por el MEC en 1994.

En este documento sobre *Centros educativos y calidad de la enseñanza* se contemplan 77 medidas para mejorar la calidad del sistema educativo, objetivo éste fijado por la LOGSE y señala, entre otros, cuatro factores que tienen incidencia directa con la enseñanza:

- La autonomía de los centros docentes.
- La dirección y la participación de la comunidad educativa en el gobierno de los mismos.
- La formación y las perspectivas profesionales de los docentes.
- La evaluación del sistema educativo y la función inspectora.

Podemos observar como la LOGSE apuesta por mejorar la calidad del sistema educativo.

Sin embargo, los centros educativos no disponen de un modelo de evaluación que, adaptado a la realidad del mismo, les permita generar una cultura de calidad. El presente programa se inicia con el propósito de desarrollar un modelo de gestión de calidad adaptado a la idiosincrasia del centro educativo, para que una vez implementado permita instaurar la cultura de la calidad. Pretendemos llegar a un modelo centrado en las necesidades del profesorado, personal no docente, familias y alumnado.

Para ello, utilizaremos de una parte, la autoevaluación, entendiendo que cada institución es la que debe decidir sobre los cambios que necesita con mayor urgencia y, por tanto, las estrategias a poner en práctica con tales fines. Cada centro educativo, en función del contexto, deberá concretar cuáles son las innovaciones más oportunas y establecer un orden entre ellas. La autoevaluación o autorrevisión interna constituye la estrategia a través de la cual las instituciones educativas detectan sus puntos fuertes y débiles y establecen programas orientados hacia la mejora (Fullan, 1986). De otra parte, utilizaremos la evaluación externa, ya que la participación en el proceso de otros evaluadores, hace que la revisión interna adquiere credibilidad en la medida que es contrastada externamente.

Técnicas de registro y análisis de la información

Los datos e información serán tratados en dos niveles:

a) Análisis cualitativos de los grupos de discusión y de los diferentes discursos de los profesores universitarios y no universitarios.
   Estos análisis se realizarán mediante los siguientes programas estadísticos: AQUAD 5.1 y NUDIST.
b) Análisis cuantitativos para conocer los índices demandados por los cuestionarios y poder establecer numéricamente la situación de determinados indicadores.
   El programa utilizado será el SPSS 9.0

Tipos de decisiones

En el proceso se deben conjugar tanto la información obtenida sobre el centro como los juicios que formulan los expertos, a través de los informes que se emitan, para a partir de ellos, establecer planes de mejora y una cultura de la evaluación entre todos los implicados en la evaluación del centro.

Las decisiones que se tomen tras detectar los puntos débiles del centro estarán orientadas a la mejora global del centro, siendo por ello decisiones consensuadas por todos los miembros de la comunidad educativa.

Informe

Destinatarios: Profesores/as, familias, alumnado.

Estructura: Está pendiente de elaborarse el informe por estar en la última fase de evaluación.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 653-658

UNIVERSIDAD DE JAÉN

Juan Ruiz Carrascosa

TÍTULO DEL PROGRAMA: **PROGRAMA DE INICIACIÓN A LA DOCENCIA UNIVERSITARIA**

RESPONSABLE DEL PROGRAMA

Secretariado de Metodología del Gabinete de Calidad de la Universidad de Jaén.

RESPONSABLE DE LA EVALUACLÓN

Director del curso.

OBJETO EVALUADO

El programa de iniciación a la docencia universitaria en la Universidad de Jaén (UJA) es una de las actuaciones contempladas en el Plan Global para la mejora de la Calidad de esta Universidad. El programa se desarrolla durante seis meses a través de un curso, desde febrero hasta junio.

Este programa tiene como destinatario al profesorado no permanente, con dedicación a tiempo completo, de reciente incorporación a la Universidad. Durante cuatro cursos han podido realizarlo 127 profesores de los diferentes departamentos.

A continuación se exponen los objetivos, contenidos y metodología propuesta.

Objetivos:

— Comprender la importancia del conocimiento de las características psicopedagógicas de los alumnos universitarios.

— Conocer, emplear y valorar métodos y técnicas útiles en el proceso de enseñanza-aprendizaje universitario.
— Considerar la importancia de la planificación de la actividad docente.
— Familiarizarse con los aspectos generales relacionados con la investigación.
— Aplicar los conocimientos adquiridos en la realización de un trabajo práctico de planificación docente.

Contenidos:

— Características psicopedagógicas de los alumnos universitarios.
— El aprendizaje: modelos, tipos, estrategias cognitivas y metacognitivas.
— La metodología didáctica en la enseñanza universitaria: modelos de enseñanza, procesos de interacción, la función tutorial, sentido y función de las prácticas.
— Evaluación de los procesos de enseñanza-aprendizaje.
— Aspectos relacionados con la investigación.
— Recursos tecnológicos para la docencia universitaria.
— Planificación de la actividad docente.

Metodología:

El curso, con carácter semipresencial, se desarrolla en dos fases. En la primera se imparten los contenidos teóricos del programa, proponiéndose la realización de un trabajo práctico sobre dichos contenidos (planificación de una asignatura). El trabajo, que tiene un carácter obligatorio, se realiza durante la segunda fase, no presencial, con el apoyo del tutor que en cada departamento se determina, finalizando con la revisión del mismo al acabar el curso.

OBJETIVOS DE LA EVALUACLÓN

Valorar a partir de criterios preespecificados la calidad y los logros del programa para tomar decisiones de mejora tanto del programa como del personal implicado.

FUNCIONES

La primera función es Normativa. Se evalúa para promover la mejora del programa, del personal implicado de forma directa, y, de modo indirecto, del colectivo en que se encuentra inmerso. Una segunda función, que afecta a los alumnos, es la de certificación en el ámbito de las actividades de formación que se realizan en la universidad respecto al profesorado.

PARADIGMA

No se ha optado por un único paradigma. Se parte del principio de complementariedad metodológica. Consecuentemente se plantea la utilización de diversas fuentes

productoras de infonnación, así como el uso de técnicas variadas y complementarias para su recogida, elaboración y tratamiento.

MODELO DE EVALUACLÓN

Considerando las características del programa objeto de evaluación, se ha seguido, en lo fundamental, la propuesta metodológica para la evaluación de programas educativos de Pérez Juste (1995)*.

METODOLOGÍA DE LA EVALUACIÓN

Se ha respetado la esencia del modelo adoptado realizando una evaluación referida a los tres momentos (al principio, durante el proceso y al final del programa). Por las características concretas del programa y sus destinatarios se ha realizado una adaptación de las dimensiones y objetos de atención contemplados en la propuesta metodológica seguida.

Primer momento

La información correspondiente al primer momento ha sido recabada a partir de diversas fuentes productoras de datos: los alumnos del curso han puesto de manifiesto sus necesidades de formación, su experiencia, expectativas y posibilidades de implicación, a través de un cuestionario.

Los expertos en formación del profesorado han aportado su valoración sobre la viabilidad y relevancia del programa, así como sobre los objetivos y contenidos planteados.

Los profesores del curso han sido consultados sobre la relación de los contenidos a impartir, las actividades a realizar y los medios y recursos a emplear.

Segundo momento

En la evaluación del desarrollo del programa han intervenido tanto los profesores como los alumnos. Los primeros han proporcionado información relevante mediante una entrevista en profundidad. Los alumnos han proporcionado información por dos vías complementarias: por medio de un cuestionario específico dirigido a todos ellos y cumplimentado al finalizar la fase presencial y no presencial, y mediante entrevistas en profundidad para analizar cómo han percibido el desarrollo del curso una vez realizado éste.

---

* Pérez Juste, R. (1995) Metodología para la evaluación de programas educativos, en A. Medina y L.M. Villar. Evaluación de Programas Educativos, Centros y Profesores. Madrid: Universitas

630

Tercer momento

En este momento, en primer lugar, se han constatado los niveles de logro en base a los criterios preespecificados, efectuándose, así mismo, una valoración global de la calidad del curso. En segundo lugar se ha cumplido con la tarea de certificación respecto a los alumnos que han desarrollado el programa, así como con los profesores tutores colaboradores.

Técnicas de recogida de información

A fin de disponer de una información fiable, relevante y válida para la evaluación y mejora del programa, se han utilizado diversas técnicas e instrumentos en cada uno de los tres momentos:

— Cuestionario para análisis de necesidades a los alumnos y cuestionario para expertos sobre análisis y valoración del programa en el primer momento.
— Entrevistas en profundidad con los profesores y alumnos en el segundo momento.
— Cuestionario específico dirigido a todos los alumnos del curso en el tercer momento.

Criterios utilizados para valorar la información

Teniendo en cuenta los objetos de evaluación considerados en los diferentes momentos del programa, los criterios han sido variados tal como se contemplan en los cuadros 1 y 2.

Técnicas de registro y análisis de la información

A partir de los datos obtenidos en las entrevistas y los cuestionarios se ha procedido a su registro y análisis, tanto cuantitativo como cualitativo, para su estudio, integración en el informe y toma de decisiones.

Tipos de decisiones

Considerando la función de mejora asignada al programa y a las características del mismo, las decisiones tomadas básicamente son de dos tipos:

— Decisiones de mejora tomadas durante el desarrollo del programa, que afectan fundamentalmente al desarrollo de las actividades propuestas y al incremento de los materiales, a partir de las solicitudes de los propios alumnos.
— Decisiones de mejora a partir del análisis de los resultados del programa una vez realizado. Estas decisiones afectan a la continuidad de programa en fases posteriores.

631

Informe

El informe tiene en principio dos destinatarios distintos: en primer lugar el Equipo de Gobierno de la Universidad, para su conocimiento y hacer posible, a la vez, la toma de decisiones. El segundo grupo de destinatarios está formado por todos los implicados en el desarrollo del programa, profesores y alumnos, haciéndose extensivo al resto de la comunidad universitaria interesada en su conocimiento.

En el informe se contempla, junto al programa del curso desarrollado, los resultados obtenidos a través de las distintas técnicas en instrumentos utilizados, así como las propuestas de mejora efectuadas.

CUADRO 1
CUESTIONARIO A EXPERTOS

EVALUACIÓN DEL PROGRAMA DE INICIACIÓN A LA DOCENCIA UNIVERSITARIA
CUESTIONARIO A EXPERTOS
ANÁLISIS Y VALORACIÓN DEL CONTENIDO DEL PROGRAMA

A fin de disponer de información fiable, relevante y válida para la mejora del Programa de Iniciación a la Docencia Universitaria, le rogamos que exprese su valoración sobre los siguientes elementos del mismo. Para ello rodee con un círculo las palabras SI o NO según convenga en cada caso, y valore, entre 1 (mínimo) y 5 (máximo) los aspectos relacionados.

Gracias por su colaboración

1.	LOS OBJETIVOS						
	1.1. Ausencia de objetivos importantes	SI	NO				
	1.2. Presencia de objetivos irrelevantes	SI	NO				
	En caso de valoración desfavorable, le agradeceríamos que indique al dorso los objetivos que faltan y/o sobran.						
	1.3. Relevancia de los objetivos planteados	1	2	3	4	5	
	1.4. Congruencia con las:						
	1.4.1. Características de los destinatarios	1	2	3	4	5	
	1.4.2. Necesidades de formación	1	2	3	4	5	
	1.5. Utilidad para la actividad profesional	1	2	3	4	5	
2.	EL CONTENIDO DEL PROGRAMA						
	2.1. Relevancia	1	2	3	4	5	
	2.2. Tratamiento						
	2.2.1. Inadecuado-adecuado	1	2	3	4	5	
	2.2.2. Desequilibrado-equilibrado	1	2	3	4	5	
	2.2.3. Desfasado-actualizado	1	2	3	4	5	
	2.2.4 Insuficiente-suficiente	1	2	3	4	5	
	2.3. Utilidad para el desarrollo profesional	1	2	3	4	5	
3.	METODOLOGÍA PROPUESTA						
	3.1. Congruencia con la modalidad de formación	1	2	3	4	5	
	3.2. Utilidad para el desarrollo del programa	1	2	3	4	5	

CUADRO 2
CONTENIDO DE LA EVALUACIÓN REALIZADA POR LOS ALUMNOS

OBJETIVOS
Los criterios se centran en la relevancia, la congruencia y la utilidad

CONTENIDOS
El criterio en este punto fue el de calidad, concretado en otros como la claridad, la estructuración, la utilidad, el valor orientador y la aplicabilidad.

ORGANIZACIÓN
1. Lugar
2. Fechas
3. Horario
4. Duración del curso
5. Condiciones del lugar
6. Número de participantes
7. Duración de las sesiones

El criterio para su valoración fue el de *adecuación*.

PONENTES
1. Capacidad de motivación
2. Capacidad expositiva
3. Facilidad para mantener relaciones interpersonales

El criterio utilizado para su valoración fue el de *adecuación*.

METODOLOGÍA
1. Presentación de teoría y conceptos
2. Presentación de material de apoyo
3. Propuestas para la práctica
4. Actividades

Los criterios fundamentales tienen que ver con la *adecuación*, la *congruencia* con los objetivos y la *claridad* de las actividades propuestas.

AMBIENTE
1. Participantes
2. Relaciones interpersonales

Los criterios fundamentales se centran en el *interés*, la *motivación* y la *adecuación* de las relaciones entre los participantes.

RECURSOS MATERIALES
El criterio en este punto fue el de su *calidad*, concretado en otros como la *adecuación*, la relevancia, la *claridad*, el grado de *utilidad* y la *aplicabilidad* práctica.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 659-664

UNIVERSIDAD DE MÁLAGA

Juan Carlos Tejar y José Serrano Angulo

TÍTULO DEL PROGRAMA: **PROGRAMA DE FORMACIÓN DEL PROFESORADO UNIVERSITARIO**

RESPONSABLE DEL PROGRAMA

Instituto de Ciencias de la Educación de la Universidad de Málaga* y Vicerrectorado de Ordenación Académica de la Universidad de Málaga

RESPONSABLE DE LA EVALUACIÓN

Directores del proyecto: Juan Carlos Tójar y José Serrano.

Objeto evaluado

Formación permanente del profesorado universitario en la Universidad de Málaga. Actividades de heteroformación (cursos, talleres, conferencias,...), experiencias de innovación docente, otras actividades de formación.

OBJETIVOS DE LA EVALUACIÓN

La puesta en marcha, el desarrollo y los resultados de esta evaluación deben contribuir a la *mejora y el perfeccionamiento de la Educación Superior* como objetivo marco. Este principio se desarrolla en una serie de objetivos generales y específicos. Los primeros se refieren a (1) la determinación y concreción de los medios y recursos necesarios para que el profesorado universitario desarrolle procesos de mejora y

---

* Hasta su desaparición en 1999.

perfeccionamiento en cuestiones docentes, y (2) la comprensión de los procesos de innovación educativa, en sí mismos, que permita un análisis de los elementos principales que favorecen la autoformación, interformación y/o ecoformación del profesorado.

Los objetivos específicos se encuentran relacionados con el rol profesional de los profesores (Scriven, 1988, 1994) y se pueden concretar como sigue:

a) Delimitar y estudiar las *competencias instruccionales* (destrezas comunicativas, de gestión, de programación y desarrollo) que pueden favorecer los procesos de innovación.
b) Comprender y analizar en profundidad las *competencias de evaluación* (conocimientos sobre evaluación de los estudiantes, destrezas de construcción y administración de pruebas, evaluación del proceso y del resultado, rendimiento de los alumnos) promovidas en la innovación.
c) Reunir elementos que permitan una reflexión crítica sobre diversas cuestiones relativas a la *Profesionalidad* (ética, actitud, mejora, servicio...).

Función: Formativa

Paradigma: Ecléctico

MODELO

Contexto de implantación del programa, diseño y planificación de las actividades, proceso de implantación y resultados y cambios producidos por el mismo (previstos y no previstos en los objetivos del programa).

METODOLOGÍA DE LA EVALUACIÓN

Diseño

Para abordar un estudio riguroso y exhaustivo de los procesos de innovación educativa en la universidad, como en otros ámbitos de la educación, resulta imprescindible en la actualidad, si no se quiere pecar de simplista, plantear un doble diseño de investigación. Es por esto que planteamos, por un lado un diseño descriptivo (enfoque metodológico empírico-analítico), a partir de ahora diseño (A), y por otro un diseño multicaso de estudio de casos (enfoque metodológico humanístico-interpretativo) a partir de ahora diseño (B). Para ambos diseños se van a describir a continuación los elementos principales previstos para la investigación.

El diseño (A), por su naturaleza, permite el enunciado de un conjunto de hipótesis específicas amplio, si bien todas ellas se pueden reunir bajo una más general, cuyo enunciado haría referencia a la *eficacia de la innovación en el desarrollo profesional docente del profesorado* (hip. 1), y por ende, una *mejora de la calidad de la enseñanza a raíz de los procesos de innovación emprendidos y desarrollados* (hip. 2).

Ambas hipótesis generales (1 y 2), que se encuentran como se ha señalado relacionadas, se pueden desarrollar en un gran número de hipótesis específicas, tales como: *la innovación mejora las competencias instruccionales del profesorado implicado en las mismas* (hip. 1.1), concretamente *las destrezas comunicativas* (hip. 1.1.1), *las destrezas de gestión* (hip. 1.1.2), *de programación* (hip. 1.1.3) y *desarrollo de esa programación* (hip. 1.1.4). Así mismo se habría de incluir la *transformación de las competencias de evaluación del profesorado implicado en innovaciones* (hip. 1.2), en concreto, *un mayor y mejor conocimiento sobre evaluación de los estudiantes* (hip. 1.2.1), *mejora de las destrezas de construcción y administración de pruebas* (hip. 1.2.2), *inclusión de aspectos tales como la evaluación del proceso junto al tradicional resultado* (hip. 1.2.3), *modificación del rendimiento de los alumnos* (hip. 1.2.4). El profesorado implicado en la innovación debe además modificar diversas actitudes que conforman las concepciones que tengan de su profesión (hip. 1.3), autoestima (hip. 1.3.1), valores éticos (hip. 1.3.2), servicio público (hip. 1.3.3), actitud de mejora y autosuperación (hip. 1.3.4).

Otras hipótesis relacionadas con las anteriores, dentro de este diseño (A) son las siguientes: *grado de cumplimiento de los objetivos de la innovación* (hip. 2.1); *permanencia en el tiempo de los cambios producidos* (hip. 2.2); *modificaciones habidas dentro de la propia innovación* (hip. 2.3); *efectos colaterales, previstos o no* (hip. 2.4), *ampliación de fines* (hip. 2.5); *difusión de la innovación en el entorno cercano* (hip. 2.6), *institucionalización de la innovación* (hip. 2.6.1); *impacto en el estudiante* (hip. 2.1).

El diseño (B) por su naturaleza impide ser tan explícito en la concreción de las hipótesis, siendo el enunciado de éstas más un medio y un fin que el inicio del proceso investigador. En efecto, no se trata de conocer cómo el profesorado lleva a la práctica las diversas innovaciones proyectadas como el comprender cómo estas innovaciones son representadas y percibidas desde los propios participantes en la innovación. En este sentido, los estudios de casos previstos seguirían el procedimiento característico del método etnográfico, que supone un *proceso interactivo* entre las diversas fases de selección de informantes y fuentes de información, recogida y análisis de la información y elaboración de categorías/dimensiones e *hipótesis* que han de contrastarse de nuevo con las fuentes y con nuevos análisis en un proceso recurrente.

Las variables implicadas en el diseño (A) vienen articuladas por la hipótesis general 1, y por tanto con la relación recíproca entre el desarrollo de la innovación y el desarrollo profesional docente del profesorado participante en la misma. Todas ellas pueden englobarse en tres generales: a) *competencias instruccionales*, (b) *competencias de evaluación* y c) *profesionalidad*. La primera (a) hace referencia a las destrezas comunicativas en el aula y con el resto de los compañeros, a la gestión e interacción en el aula, a la programación y desarrollo de las materias que han de enseñarse. La dimensión b) implica conocimientos sobre evaluación de los estudiantes, destrezas de construcción y administración de pruebas de conocimientos y procedimientos, evaluación del proceso y del resultado, rendimiento de los alumnos. La dimensión c) involucra aspectos éticos, de actitud hacia la profesión, implicación, de servicio público, de superación, autoestima, etc.

En el diseño (B) las dimensiones (o macrocategorías) de análisis de partida pueden ser los cambios en a) las *concepciones curriculares*, b) las *condiciones que facilitan el cambio* y c) las aportaciones y exigencias que la innovación implica a cada profesor participante en una innovación. En a) se trata de identificar cambios en la concepción y la práctica de la enseñanza, aprendizaje y la evaluación. La dimensión b) recoge aspectos que inciden en la organización de la innovación: modelo de innovación, negociación sobre la innovación, interacción profesor-alumno, apoyo institucional y seguimiento de la innovación. En c) se pueden configurar, entre otros, los siguientes aspectos: mejora de la autoestima profesional, mejora de relaciones entre el profesorado, apertura a nuevos valores profesionales (reflexión crítica, investigación en el aula, formación permanente...).

Técnicas de recogida de información

Tanto en (A) como en (B) se utilizarón similares fuentes de información en principio, esto es, principalmente toda la documentación disponible en la Universidad de Málaga sobre proyectos de innovación desarrollados en el pasado y en proceso en la actualidad (convocatorias, proyectos iniciales, informes, memorias, publicaciones, ...), las opiniones de los responsables institucionales, del profesorado y del alumnado implicado. Sin embargo, existen claras diferencias derivadas de los instrumentos y técnicas de recogida de la información.

Como técnicas de recogida de datos en (A) se aplicaron *cuestionarios* a una muestra aleatoria de profesores de la Universidad de Málaga y a una muestra aleatoria de alumnos implicados, y *entrevistas estructuradas* a los coordinadores de los equipos de profesores y a los responsables institucionales. Sobre los documentos recopilados se aplicaron diversas *hojas de registro* (planillas) de las que se obtuvo información sobre las variables relevantes.

Las técnicas de recogida de información previstas en (B) son el *diario de campo*, lógicamente, la *observación participante*, la *entrevista en profundidad*, la *fotografía* de escenarios relevantes y la aplicación de *cuestionarios abiertos* de manera limitada. Previa negociación se fijó, en cada caso, el nivel de participación permitido en según qué circunstancias (reuniones de coordinación, seguimiento...). Sobre los documentos referidos anteriormente se realizarón *análisis de contenidos*.

Criterios utilizados para valorar la información

En general la suficiencia, la viabilidad y la evaluabilidad de la información. En particular, para el diseño (a) la consistencia de los juicios y la contrastación de las informaciones, y para (b) la credibilidad, la confirmabilidad, la dependencia y la transferibilidad de la información.

Referencias

Progreso del propio programa.

637

Técnicas de registro y análisis de la información

Para analizar los datos obtenidos a partir del diseño (A) se realizaron sendos análisis descriptivos univariante y bivariante, fundamentalmente análisis de frecuencias y porcentajes, tablas de contingencia y ji-cuadrado. También fue muy útil realizar un análisis multivariante de reducción de datos para obtener criterios para clasificar, agrupar y organizar los datos representando las características más relevantes de los mismos. Asimismo, se realizaron contraste de hipótesis para señalar las diferencias entre las respuestas de los profesores participantes en proyectos de innovación y/o talleres y las respuestas de los profesores no participantes.

La intención interpretativa del diseño (B) de la investigación dota al análisis de datos de una característica peculiar en el sentido de que no lo circunscribe a una fase o periodo determinado a diferencia de lo que ocurre en el otro diseño. Como ya se ha indicado anteriormente, el análisis de datos se hizo de manera *recurrente*, esto es, se produjo desde momentos anteriores a la propia recogida de datos, durante ella misma y, cómo no, tras recoger la información relevante.

Durante el análisis cualitativo se realizaron diversas actividades y construcciones según una escala creciente de abstracción y generalización, como es: un *análisis especulativo*, la *clasificación y categorización*, la *formación de conceptos*, los *modelos y tipologías*, de forma que, progresivamente, los datos iban tomando cuerpo de *teoría*.

Como estrategias o procedimientos generales se utilizaron la *inducción analítica* y las *comparaciones constantes*, como técnicas inductivas para la generación de teoría; el *análisis tipológico*, que debe permitir de manera flexible la consideración o eliminación de categorías según su relevancia; y la *enumeración*, cuando era posible categorizar la información con la claridad suficiente.

Todos estos procesos de teorización genéricos y procedimientos analíticos mencionados anteriormente constituyen verdaderos instrumentos conceptuales para la elaboración de los análisis. El proceso de análisis, que se extendió incluso hasta la integración e interpretación de los resultados, se ayudó de la elaboración de gráficas descriptivas (esquemas contextuales y de evaluación de una situación), gráficas explicativas (diagramas de dispersión, causales y esquemas de flujo del estado de cualquier acontecimiento), matrices descriptivas (v. g. ordenadas o meta-matrices) y explicativas (v. g. de efectos).

La triangulación se utilizó para analizar aspectos fundamentales de las experiencias de innovación, debido a la posibilidad de fluidez y contraste de la comunicación que se transmitio desde los diversos ángulos o perspectivas implicadas

Para integrar la información se usó análisis de la red de datos cualitativos, análisis de discursos, análisis fenomenológico de datos de las entrevistas y el análisis de reuniones situacionales.

Tipos de decisiones

Cada fuente de información, y todos los análisis y valoraciones que se hicieron a partir de cada una, están a disposición de los correspondientes implicados para garan-

tizar la transparencia del proceso y que las conclusiones del trabajo sean compartidas y ayuden a encontrar vías de solución a las necesidades allí donde se originen. Además, las diversas autoridades educativas (universidades, consejería y ministerio) son informadas para que tomen conciencia de los resultados del estudio y los puedan tener en cuenta a la hora de la planificación y gestión de los recursos para la enseñanza en la Universidad.

Informe

Los informes de los estudios de caso tienen estructura y amplitud variable, y tienen como destinatario principal los participantes de cada «caso».

El informe global tiene como destino el Centro de Documentación e Investigación Educativa (CIDE), y tiene la estructura de acuerdo con la convocatoria publicada en la orden de 23 de septiembre de 1997 por la que se convoca el concurso nacional para otorgar ayudas a la investigación educativa. (B.O.E. núm. 243, de 10 de octubre de 1997). Una versión resumida de este informe global será entregada al resto de autoridades educativas con el permiso del organismo convocante de las ayudas.

639

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 665-670

UNIVERSIDAD DE MURCIA

Pilar Martínez Clares y María Paz García Sanz
Departamento de Métodos de Investigación y Diagnóstico en Educación

PRIMER INFORME

TITULO DEL PROGRAMA: **EL DESARROLLO PERSONAL Y SOCIAL: EL AUTOCONCEPTO**

RESPONSABLE DEL PROGRAMA:

La orientadora.

RESPONSABLE DE LA EVALUACIÓN

La orientadora-investigadora.

OBJETO EVALUADO

Se trata de un programa de desarrollo personal y social instructivo que proporciona estrategias y habilidades que fomentan el desarrollo de un positivo autoconcepto, destinado a los alumnos del tercer ciclo de Primaria y de Educación Secundaria Obligatoria.

OBJETIVOS DE LA EVALUACIÓN

— Controlar y analizar la evaluación de la viabilidad del programa, a través de un estudio de la calidad técnica del mismo y su adecuación y adaptación al contexto.
— Realizar la evaluabilidad o validación del programa a través de una serie de criterios e indicadores de evaluación.

640

— Analizar la evaluación de proceso del programa aplicado por medio de la evaluación de la implementación y de la evaluación de desarrollo.
— Analizar y estudiar los efectos del programa adaptado y diseñado específicamente de autoconcepto a través de la evaluación de la eficacia y efectividad.
— Realizar una valoración general del programa contrastando las modificaciones o cambios producidas en las variables de estudio.

Función: Se trata de una evaluación formativa y sumativa con una finalidad preventiva y de desarrollo de todo el proceso.

Paradigma: Trabajamos desde un paradigma ecléctico; nuestro enfoque se basa fundamentalmente en un modelo tecnológico: dependiendo de las fases de la evaluación nos situamos más en postura más cualitativa o más cuantitativa.

Modelo: Modelo de evaluación de programas de Hernández Fernández, J. y Martínez Clares, P. (1996). Modelo que consta de cinco fases:

1ª Fase: Supuestos teóricos o previos
2ª Fase: Viabilidad de la evaluación
3ª Fase: Evaluabilidad o validación de la evaluación.
4ª Fase: Evaluación del proceso: evaluación de la implementación y evaluación de desarrollo.
5ª Fase: Evaluación de la eficacia.

METODOLOGÍA

Diseño: Hemos utilizado un diseño cuasi-experimental-descriptivo, ya que tratamos de establecer relaciones entre el programa y sus posibles logros y/o resultados, y porque nos detenemos en describir los procesos que ocurren en la aplicación del programa en sus distintas fases. Se trata de un diseño cuasi experimental-descriptivo con un diseño de grupo de control no equivalente.

Técnicas de recogida de datos: Para la exploración inicial del autoconcepto utilizamos el cuestionario de Autoconcepto de Musitu, García, y Gutierrez (1991) AFA: Autoconcepto. Forma-A. También se utilizó un diario de investigación y una escala de estimación de observación de las distintas sesiones. Igualmente se realizaron entrevistas con los tutores de los distintos grupos de alumnos y, al finalizar cada sesión, se procedía a una mini-asamblea para obtener información directa durante todo el proceso. Asimismo, otro instrumento de recogida de datos fueron los propios materiales de los alumnos/as derivados de la realización del programa. Al finalizar el programa, se procedió a la cumplimentación de un cuestionario de evaluación por parte del alumnado y los tutores.

Técnicas de registro y análisis de la información: todos los datos recogidos en el estudio o trabajo de campo fueron introducidos en una base de datos (dBase III Plus) y para el análisis estadístico de resultados se utilizó el paquete estadístico SPSS/PC+.

641

Se fueron realizando diferentes tipos de análisis en función del tipo de variable que relacionábamos con el autoconcepto. Fundamentalmente se llevó a cabo un estudio descriptivo para poder realizar un estudio analítico y terminamos con un estudio comparativo entre los grupos.

Informe: Concluimos este trabajo elaborando un doble informe: por una parte se realizó un informe para el centro, destinado al claustro, al consejo escolar y, principalmente, a los profesores tutores implicados en el mismo. Por otra, se realizó una tesis doctoral, con una explicitación detallada de todas las fases de la evaluación del programa y las conclusiones.

BIBLIOGRAFÍA

Hernández Fernández, J. y Martínez Clares, P. (1996). Propuesta metodológica para evaluar programas de orientación educativa. *Revista Electrónica de Investigación y Evaluación Educativa*. (RELIEVE). Vol. II, nº. 1. 2-3.

Martínez Clares, P. (1996). *Diseño y Evaluación de un Programa de Desarrollo Personal y Social: El Autoconcepto*. Tesis Doctoral. Inédita. Facultad de Educación: Universidad de Murcia.

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **PROYECTO CURRICULAR DE EDUCACIÓN SECUNDARIA OBLIGATORIA**

RESPONSABLES DEL PROGRAMA

Profesores y profesoras que imparten docencia en la etapa de Educación Secundaria Obligatoria.

RESPONSABLES DE LA EVALUACIÓN

Todos los profesores y profesoras de Educación Secundaria Obligatoria del Instituto, así como —con una mayor implicación—, la Comisión de Coordinación Pedagógica del Centro y una investigadora externa, que actúa como asesora, compartiendo responsabilidades con el resto de agentes evaluadores.

642

OBJETO EVALUADO

Proyecto Curricular de Educación Secundaria Obligatoria.

OBJETIVOS DE LA EVALUACIÓN:

1. Permitir un amplio conocimiento de los elementos que configuran el Proyecto Curricular de Etapa (PCE) así como una reflexión conjunta del desarrollo y resultados del mismo.
2. Reflexionar sobre la praxis evaluativa actual y adoptar un sistema de evaluación que responda a un proceso funcional de diálogo, consenso y mejora.
3. Impulsar la autoevaluación en el Instituto con el fin de mejorar la calidad de los procesos de enseñanza-aprendizaje que en él se imparten.
4. Difundir una cultura evaluadora dentro del instituto, de modo que la evaluación quede enmarcada dentro de las propias actividades de la acción docente, haciendo latente la percepción de que la misma es conveniente, necesaria y favorable para el propio Centro.
5. Potenciar la evaluación como elemento dinamizador de los procesos educativos, favorecedor de acciones coherentes y fundamentadas y de actuaciones colegiadas y comprometidas, capaces de superar acciones improvisadas e individualistas.
6. Considerar la evaluación del PCE como un proceso inacabado, permanentemente revisable y abierto a reajustes consensuados.
7. Ofrecer al Instituto procedimientos e instrumentos de evaluación que faciliten la evaluación interna del propio PCE.
8. Reflexionar sobre la adecuación de los objetivos generales de Etapa y los de cada una de las áreas y materias y evaluar su adecuación al contexto del Instituto.
9. Estudiar la selección, distribución y secuenciación de contenidos inter e intraciclos e introducir las modificaciones pertinentes según vayan surgiendo.
10. Valorar la idoneidad de la metodología explicitada en el PCE e ir estableciendo los cambios oportunos en este ámbito, así como en lo que respecta a las decisiones sobre agrupamientos, organización de tiempos y espacios y de los materiales curriculares y didácticos empleados.
11. Reflexionar sobre la validez de las estrategias y procedimientos de evaluación y establecer la correspondiente toma de decisiones que conduzcan hacia una mejora de la evaluación del aprendizaje de los alumnos, la práctica docente y el propio PCE.
12. Analizar las actividades de orientación educativa y profesional explicitadas en el PCE y evaluar su idoneidad.
13. Valorar la oferta de materias optativas realizada por el Instituto y comprobar su adecuación a las necesidades educativas de los alumnos.
14. Evaluar la validez de los criterios aplicados en la adaptación del currículo para alumnos con necesidades educativas especiales, así como la efectividad de los programas de diversificación curricular adoptados en el Instituto.

15. Fomentar la metaevaluación como proceso para apreciar si el método y las técnicas empleadas han sido adecuadas, si los logros responden a los objetivos propuestos y si el proceso de evaluación ha respetado las pautas éticas respecto a los evaluados y a los evaluadores.

Funciones: Función formativa, la cual ha permitido conocer en profundidad, comprender y mejorar el PCE; función sumativa, que ha logrado sintetizar, explicar y valorar los resultados de la aplicación del PCE; por último, función administrativa, la cual ha permitido responder a las prescripciones exigidas por el MEC (Hernández Pina, 1995).

Paradigma: Enmarcamos la evaluación del PCE dentro del paradigma orientado hacia la toma de decisiones y el cambio, en donde hemos dado una especial relevancia a los aspectos cualitativos, aunque no hemos descartado el empleo de técnicas cuantitativas.

Modelo: Desarrollado por García Sanz (1999) y Hernández Pina y García Sanz (en prensa).

METODOLOGÍA DE LA EVALUACIÓN

- **Muestra:** 53 profesores que impartieron docencia en ESO en un Instituto durante el curso académico 1997-98.
- **Técnicas e instrumentos:** Para la recogida de información contamos con los siguientes tipos de técnicas: observación participante, entrevistas, grupos de discusión, análisis de contenido, supervisión clínica, triangulación y un cuestionario para evaluar el PCE.
- **Procedimiento:** La evaluación del PCE se ha llevado a cabo mediante procedimientos propios de la investigación cooperativa, distinguiéndose tres momentos: una fase inicial, que ha incluido la realización de los preparativos, los primeros contactos con el Centro y la elaboración provisional de un cuestionario base para evaluar el PCE; una fase de desarrollo, en la que se ha llevado a cabo la evaluación inicial, procesual y sumativa del PC; y una fase final, que ha incluido un análisis cualitativo y cuantitativo de los datos y la metaevaluación del proceso.
- **Tipos de decisiones:** Las decisiones tomadas se han traducido en el establecimiento de una serie de propuestas para mejorar el PCE y su puesta en práctica, a realizar a corto y medio plazo.

Informe de evaluación: Se han realizado dos tipos de informes: Por una parte, uno de corta extensión, incluido en la memoria final del curso, por lo tanto, dirigido al profesorado y a la Administración. Por otra parte, un informe mucho más extenso —Tesis Doctoral (García Sanz, 1999)—, dirigido a la comunidad científica.

644

BIBLIOGRAFÍA

García Sanz, M.P. (1999). *Evaluación del Proyecto Curricular de Educación Secundaria Obligatoria en un Instituto de Educación Secundaria. Un caso de Investigación Cooperativa.* Tesis Doctoral Inédita. Facultad de Educación. Universidad de Murcia.

Hernández Pina, F. (1995). *Bases Metodológicas de la Investigación Educativa.* Murcia: DM.

Hernández Pina, F. y García Sanz, M.P. (2001). *Evaluación del Proyecto Curricular de Educación Secundaria Obligatoria.* Madrid: La Muralla.

645

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 671-679

UNIVERSIDAD DE NAVARRA

Charo Reparaz
Departamento de Educación

PRIMER INFORME

TÍTULO DEL PROGRAMA: **PROYECTO EGO-ETNO (EGOCENTRISMO-ETNOCENTRISMO) DE INVESTIGACIÓN-ACCIÓN Y FORMACIÓN PEDAGÓGICA EN EDUCACIÓN INTERCULTURAL**

Dirigido por un equipo investigador del *Instituto Regionale di Ricerca Sperimentazione Aggiornamiento Educativi* (I.R.R.S.A.E.-Lombardía) de Italia. Investigador principal Cesare Scurati. Los países participantes fueron: Italia, Francia, Gran Bretaña, Alemania y España. Al Departamento de Educación de la Universidad de Navarra le fue confiada la realización y supervisión de la investigación en España.

RESPONSABLE DEL PROGRAMA

Supervisora: Mª del Coro Molinos Tejada
Tutora: Mª Luisa Durán Los Arcos

RESPONSABLE DE LA EVALUACIÓN

Equipo investigador del *Instituto Regionale di Ricerca Sperimentazione Aggiornamiento Educativi* (I.R.R.S.A.E.-Lombardía) de Italia.

OBJETO EVALUADO

Programa de Educación Intercultural.

646

OBJETIVOS DE LA EVALUACIÓN

Comprobar los cambios de actitudes en los alumnos, tras una doble intervención con metodología de tipo afectivo-cognitivo clínico aplicada a la discusión a) de experiencias personales del alumno, y b) de situaciones descritas en materiales didácticos, específicos para la educación intercultural.

Función: Inicial, diagnóstica y sumativa.

Paradigma: Ecléctico:
— Tests estandarizados: B.A.S. y P.F.S.E.
— Cuestionario Proyectivo basado en imágenes: *test Ego-Etno.*
(Elaborado por el equipo italiano del IRRSAE)

Modelo: Evaluación Test-Retest.

METODOLOGÍA DE LA EVALUACIÓN

Diseño:

Se evaluó la situación de entrada de los alumnos en relación a sus actitudes y habilidades sociales y el cambio experimentado: a) tras una intervención con metodología de tipo afectivo-cognitiva; y b) tras una intervención de tipo didáctico-curricular combinada con la metodología afectivo-cognitiva.

Técnicas de recogida de la información:

— B.A.S.-3 —Batería de socialización— Silva Moreno, F. y Martorell Pallás, M. C., Ed. TEA.
— P.F.S.E. —Cuestionario de actitudes— Nivel elemental, Yuste Hernánz, C., Ed. CEPE.
— Cuestionario Proyectivo basado en imágenes: *Test Ego-Etno* (Elaborado por el equipo italiano del IRRSAE).
— Entrevistas con los profesores-investigadores.
— Observaciones de los profesores-investigadores.
— Diarios de clase de los profesores-investigadores.
— Análisis de los materiales didácticos.

Criterios para valorar la información:

La calidad técnica del Programa, su adecuación y eficacia.

Referencias:

Progreso sobre el propio Programa.

Técnicas de registro:

Sesiones participativas de exposición, debate y conclusión.
Las técnicas de análisis propias de los cuestionarios utilizados.
Auditoría externa de investigadores expertos en el tema.

Tipos de decisiones:

Continuidad y mejora del Programa.

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **ESTIMULACIÓN TEMPRANA: UNA MEDIDA DE ENRIQUECIMIENTO CURRICULAR EN EDUCACIÓN INFANTIL**

Equipo investigador del IRRSAE.

RESPONSABLE DEL PROGRAMA

Departamento de Educación de la Universidad de Navarra: Lic. Dña. Sonia Rivas Borrell. [Dirección Dra. Dña. Felisa Peralta López, parte teórica, y Dr. D. Ángel Sobrino Morrás, parte metodológica].

RESPONSABLE DE LA EVALUACIÓN:

Dña. Sonia Rivas Borrell con orientaciones del Dr. D: Ángel Sobrino Morrás.

OBJETO EVALUADO:

Programa de estimulación temprana desarrollado en el centro de educación infantil *Kutunbaita* (San Sebastián). Aspectos cognitivos, afectivos y de socialización, comunicativos y motrices.

Edades de los niños 2 y 3 años cumplidos (2 grupos distintos). Se trata de comprobar la evolución de los niños de 3 años, una vez que abandonan el programa después de dejar Kutumbaita en junio de 2000.

OBJETIVOS DE LA EVALUACIÓN:

Comprobar las variables que inciden en la eficacia y en la calidad del programa.

Función: Formativa. Para orientar a la dirección sobre puntos de mejora en su tarea educativa, e incidir o abandonar los puntos que se crean oportunos.

Paradigma: Paradigma cualitativo más que cuantitativo. A pesar de que se administren unas pruebas a los niños, no solamente se evaluará esta variable, sino que también se tendrán en cuenta otras como puede ser el profesorado, su ambiente familiar y educativo, e incluso la misma institución escolar, siguiendo los criterios de la perspectiva ecológica de Bronfenbrenner y el modelo transaccional de Sameroff.

Modelo: Se tendrá en cuenta el modelo CIPP de Stufflebeam.

METODOLOGÍA DE LA EVALUACIÓN

Técnicas de recogida de información:

— Observación no participante de los niños, de la actividad diaria de las profesoras en contextos académicos y no académicos, entrevistas a los profesores y a la directora.
— Intrumento utilizado con los niños: escala Haitzea, baremada en 1991 (estudio estadístico realizado por la editorial TEA) en población vasca.
— Escala Home, para evaluar la estimulación proporcionada en el ambiente familiar.

Criterios utilizados para valorar la información:

— Evaluabilidad del programa, eficacia y eficiencia.

Referencias:

— Estándares prefijados: objetivos curriculares que propone la LOGSE en los diversos documentos; población vasca estandarizada de la escala Haizea, que sirve como grupo de control.

Técnica de registro y análisis de la información:

Registro mediante un diario de campo. El propio evaluador es el que realiza la recogida de información: no se utilizan medios audiovisuales.

Tipos de decisiones:

De mejora del programa. Optimización educativa.

Informe

Para el centro educativo *Kutunbaita* (cliente que demandó que se realizara la evaluación). Extensible aunque no generalizable a otros centros educativos.

Parte experimental de una tesis doctoral.

TERCER INFORME

TÍTULO DEL PROGRAMA: **REVISIÓN DE SOFTWARE PARA DESARROLLAR EXÁMENES Y SER DISTRIBUIDOS A TRAVÉS DE REDES (*INTERNET* O *INTRANET*)**

RESPONSABLE DEL PROGRAMA

Diferentes compañías comerciales y universidades

RESPONSABLE DE LA EVALUACIÓN

Sonia Lara Ros

OBJETO EVALUADO:

El software actualmente disponible para realizar exámenes y ser distribuidos a través de redes (*Internet* o *Intranet*).

OBJETIVOS

Presentar las actuales posibilidades que las redes ofrecen para generar exámenes a través de redes.

Presentar las principales capacidades instructivas del software revisado para poder emplearlo en procesos de evaluación formativa, tanto en enseñanza presencial como no presencial.

Función: Revisión del software disponible atendiendo a diferentes criterios (instructivos, capacidades para hacer análisis estadísticos, costes y cuestiones técnicas)

Paradigma: cuantitativo y cualitativo

Modelo: CIPP

METODOLOGÍA DE LA EVALUACIÓN

Diseño

Técnicas de recogida de información:

En una primera aproximación se buscó y seleccionó el *software* que iba a ser objeto de la evaluación. Para conseguir este *software* se recurrió al citado en la investigación educativa (artículos, comunicaciones y libros), y a búsquedas en Internet.

La selección del *software* que finalmente se revisó atendió a varios criterios:

* *Software* que pudiera plantear la implantación de evaluación por sí mismo, haciéndose cargo de todos los aspectos necesarios (servidor, recogida de datos, y gestión de la información). Se descartaron los programas denominados *herramientas de autor* por no permitir recoger o almacenar los resultados de los exámenes en un servidor También se descarto el *software* para *construir cursos en el web*, por tener tan sólo un módulo dedicado a la evaluación, y por la necesidad de contar con todo el paquete para emplear el de evaluación.
* Dado que fue abundante la información encontrada, se seleccionó el *software* en función de la utilización o investigación que sobre él había realizada como primer filtro de la evaluación.

Para poder revisarlo se contó con una copia de demostración de todo el *software* para poder instalarlo, o de una cuenta para poder acceder a través del *web*. De este modo se analizó *in situ* todo el *software*. En total ser revisan 18 programas (10 de universidades y 8 comerciales).

Criterios utilizados para valorar la información:

Suficiencia, viabilidad, eficacia, eficiencia, efectividad

Referencias:

Estándares prefijados (criterial) y superioridad sobre otros programas

Técnicas de registro y análisis de la información:

Toda la información (gráfica, criterios de observación y comentarios) se recogió en una base de datos (File Maker) que permitió la posterior tabulación de la información recogida.

Tipos de decisiones:

Ofrecer los puntos fuertes y débiles de cada software. Ofrecer pautas para poder seleccionar un *software* que se adecúe a las necesidades de formación de una Institución educativa.

Informe

Dirigido a los docentes y profesionales de la educación que están interesados en la innovación a través de las Tecnología de la Información y la Comunicación. Nivel educativo preferente la universidad presencial y no presencial, así como empresas de formación a distancia.

CUARTO INFORME

TÍTULO DEL PROGRAMA: **PROGRAMA DE PREPARACIÓN AL INGRESO Y SEGUIMIENTO PSICOPEDAGÓGICO DEL NIÑO HOSPITALIZADO**

RESPONSABLE DEL PROGRAMA

Aquilino Polaino Lorente y Olga Lizasoáin Rumeu. Departamento de Educación. Universidad de Navarra.

RESPONSABLE DE LA EVALUACIÓN

Aquilino Polaino Lorente y Olga Lizasoáin Rumeu (Departamento de Educación. Universidad de Navarra), en colaboración con el personal sanitario de la planta de Pediatría de la Clínica Universitaria de Navarra.

OBJETO EVALUADO

Programa de Intervención Psicopedagógica en el ámbito de la Hospitalización Infantil.

OBJETIVOS DE LA EVALUACIÓN

El objetivo principal de esta investigación es ayudar al niño mediante un programa de preparación a la hospitalización y seguimiento psicopedagógico, a la vez que colaborar en su adaptación a la clínica. Se asume que esta preparación puede contribuir a disminuir los efectos nocivos y/o iatrógenos que tradicionalmente vienen atribuyéndose a la hospitalización infantil.

Función: Inicial diagnóstica y post aplicación del programa para comprobar su eficacia.

Paradigma: Cuantitativo

Modelo: Pretest-Postest

METODOLOGÍA DE LA EVALUACIÓN

Diseño:

Definición de la muestra. 40 niños ingresados en la planta de Pediatría de la Clínica Universitaria de Navarra, entre 8-12 años, con una enfermedad crónica o aguda que requieran una hospitalización de media duración (mínimo una semana y máximo 20 días). Los probandos se distribuyeron de forma aleatoria entre el grupo de control y el experimental.

Se trata de un diseño unifactorial de dos grupos, sobre los que se ha empleado la aleatorización como técnica de control. El N de los sujetos es igual en ambos grupos.

Técnicas de recogida de información:

Variables dependientes básicas. Totalidad de las variables medidas con escalas e instrumentos de evaluación:

— Depresión infantil
— Ansiedad estado y Ansiedad rasgo
— Autoconcepto infantil
— Habilidades sociales
— *Locus* de control

Variable independiente

Programa de Preparación a la Hospitalización y Seguimiento Psicopedagógico del niño hospitalizado (PHSP).

Referencias:

Programas del ámbito anglosajón, principalmente norteamericanos.

Técnicas de análisis de la información:

Análisis de covarianza, varianza y correlaciones.

Tipos de decisiones:

Difusión y aplicación más generalizada del programa en el entorno hospitalario.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 681-690

UNIVERSIDAD DEL PAÍS VASCO/EHU

B. M. Munarriz y J.F. Lukas
Departamento de Métodos de Investigación y Diagnóstico en Educación

PRIMER INFORME

TÍTULO DEL PROGRAMA: **EVALUACIÓN DEL PROGRAMA DE MEDIO ABIERTO**

RESPONSABLE DEL PROGRAMA

Consorcio para la Educación Compensatoria y Formación Ocupacional.

RESPONSABLE DE LA EVALUACIÓN

B. Munárriz y J.F. Lukas. Departamento M.I.D.E. - Universidad del País Vasco/EHU.

OBJETO EVALUADO

Programa de Medio Abierto desarrollado por Educadores Sociales

OBJETIVOS DE LA EVALUACIÓN

El objetivo general de la evaluación es valorar si el Programa de Intervención en Medio Abierto funciona, es útil y es necesario.

Los objetivos específicos:

* Comprobar si el Consorcio ha cumplido con sus objetivos y funciones.
* Analizar el grado de cumplimiento de los objetivos referidos a los Ayuntamientos.

* Valorar el grado de coordinación entre el equipo de educadores y otros agentes implicados en el trabajo con juventud.
* Analizar la puesta en marcha y desarrollo del Programa por parte de los educadores.

FUNCIÓN

El Consorcio, órgano que financia en su mayor parte el programa y realiza el seguimiento del mismo, estaba interesado en conocer las necesidades de la población a quien se dirige el programa y la eficacia de dicho programa. La función formativa es la que se manifiesta como la más importante desde las instituciones implicadas, ya que una de las finalidades es la de posibilitar el desarrollo y la mejora del mismo. A medida que avanzamos en la comprensión de los problemas somos conscientes de la función de control que asume los cumplimientos mínimos de coordinación entre instituciones. La función sociopolítica, que está implícita en toda evaluación y, especialmente, cuando se trata del proyecto coordinado por distintas instituciones, donde conviven políticas diferentes e incluso grados de implicación según la responsabilidad o titularidad en el programa.

Se trata de un trabajo de evaluación relacionado con las políticas públicas. Ello supone la necesidad de rendir cuentas y ofrecer información sobre los resultados obtenidos de cara a estimar la calidad de los programas ofertados con dinero público, así como, a plantear líneas de mejora.

PARADIGMA

La elección del enfoque de evaluación depende del tipo de programa a evaluar, de la finalidad del mismo e incluso del alcance del programa.

En nuestro caso, las características del programa nos afirmó en la idea de comprender desde dentro tanto las necesidades de los participantes como la política de las instituciones implicadas. Los supuestos del paradigma cualitativo se ajustaban al enfoque apropiado para la evaluación del programa.

MODELO

Las negociaciones con las instituciones responsables del programa nos fueron definiendo el modelo de evaluación más acorde al programa en cuestión. Conocer lo que las instituciones y los responsables de la puesta en marcha del programa pretendían con la evaluación, nos llevó a conocer el problema y enfocar nuestro estudio.

Dentro del enfoque cualitativo que planteamos, nos fuimos decantando por un modelo ecléctico que, partiendo del material disponible y la colaboración de los coordinadores en el estudio, con lo que supone de apertura o cooperación en la recogida de datos, podamos obtener un juicio informado sobre el programa.

De esta forma, nos acercamos a los modelos o propuestas realizadas por Stake, estudio naturalista de casos, y Stufflebeam, en cuanto a profundizar en la información para formular juicios que posibiliten la mejora.

METODOLOGÍA DE LA EVALUACIÓN

La propuesta de acercamos a modelos que plantean un visión holística del objeto de estudio, implica el planteamiento de una metodología que recoja o atienda a los supuestos que se requieren para llegar a procesos y resultados con la validez necesaria en todo trabajo de investigación. En este sentido, la metodología va a marcar el diseño de evaluación.

Diseño

La planificación de un diseño abierto, respondente, en el sentido de *feek-back* constante entre las audiencias implicadas y los evaluadores. Un diseño que permita crear/utilizar estrategias de recogida y análisis de datos, según las necesidades del proceso.

Así, desde la concepción de un diseño que se va estructurando a medida que avanza el proceso planteamos las diferentes estrategias.

Técnicas de recogida de información

La información necesaria, en cualquier proceso de evaluación, consta de una parte documental, que ya existe, y la que depende de nuestras indagaciones durante el desarrollo de la evaluación.

Así, contamos con a) documentos oficiales sobre objetivos y estructura del programa y b) memorias presentadas a las instituciones y que recogen la evaluación interna de los programas.

Las técnicas empleadas durante el proceso de evaluación fueron la observación participante, con una observadora en cada municipio donde el programa era objeto de evaluación; entrevistas a representantes de las diferentes audiencias relacionadas con el programa; encuestas a todos los sujetos participantes directos y/o indirectos en el programa. Cada una de las técnicas enfatiza más la recogida de un tipo de datos. La observación nos proporciona datos tanto del contexto como del desarrollo en cada uno de los lugares, recogiendo sus peculiaridades y el día a día en su puesta en marcha. La entrevista está centrada en cómo los políticos, asistentas sociales y educadores sociales sienten y perciben el programa y sus problemas mientras la encuesta recoge el impacto de los programas en la comunidad. Los diferentes datos van a posibilitar la triangulación a lo largo del proceso.

Criterios utilizados para valorar la información

En un primer momento, y antes de comenzar la evaluación, se establecieron unas primeras reuniones con los responsables del programa con el fin de analizar la viabilidad de la evaluación. La negociación estuvo centrada en los datos que nos podían proporcionar (documentos anteriores), colaboración con la recogida de datos durante el proceso, superando la petición de los educadores, que abogaban por una evaluación

desde la metodología de I/A, capaz de ayudar en su formación y las instituciones que planteaban un enfoque más centrado en resultados. Así mismo, el diseño empleado fue objeto de acuerdos entre las partes, al igual que la utilización final de la evaluación. De esta forma, los acuerdos de las partes implicadas clarificaron la viabilidad de la evaluación y sus limitaciones.

La comparación de la información recogida sobre los logros obtenidos con los objetivos previstos en el programa, nos define la *eficacia* del mismo, es decir, la medida del cumplimiento de los objetivos (objetivos/resultados). En el mismo sentido, las informaciones sobre los medios utilizados para la consecución de los resultados nos proporciona datos que nos hablan de la *eficiencia* de programa, y por último la comparación de los logros obtenidos con las necesidades de los consumidores dejan claro la *efectividad* del programa objeto de la evaluación.

En la primera fase de la evaluación que presentamos se precisaron los criterios cuya operatividad nos posibilitaba el análisis de la congruencia entre lo propuesto y lo real, teniendo en cuenta los objetivos, medios para conseguirlos y las necesidades de los participantes.

Referencias

Las utilizadas para la valoración del programa del medio abierto las agrupamos en torno a referencia *criterial*, basada en niveles especificados en la primera fase de la investigación por los responsables del programa, en nuestro caso quedaban concretados en: compromisos por parte de las instituciones, relaciones entre los departamentos, contribución económica, etc.; e *idiosincrática*, tomando como punto de comparación los logros, considerados por los responsables satisfactorios, del propio programa en los diferentes municipios.

Técnicas de registro y análisis de la información

Los sistemas de registro, lógicamente, están relacionados con las técnicas de recogida de información:

Sistemas de registro	Procedimiento análisis	Programas informáticos
*documentos varios: —normativas legales —memorias anuales del programa en cada municipio	Análisis de contenido	
*actas reuniones *notas de campo *entrevistas semiestructuradas	transcripción Análisis contenido: Categorización Codificación	AQUAD
*cuestionarios	Categorizado	SPSS

La *triangulación* de datos, momentos, técnicas, expertos y participantes fue utilizada a lo largo de todas las fases de la evaluación.

Tipos de decisiones

El hecho de ser una evaluación externa, encargada por los responsables del programa desarrollado, exige a la evaluación la información suficiente para poder tomar decisiones fuera del ámbito del equipo evaluador.

En nuestro caso, a partir del análisis de los datos se fue valorando el programa a nivel general y por municipios, teniendo en cuenta los criterios y referencias. De esta forma, durante el proceso, tomamos directamente decisiones de colaboración/ayuda de los evaluadores a quienes desarrollaron el programa, educadores sociales, en momentos puntuales y en aras de una implicación de las diferentes audiencias en la propia evaluación. Las posibles mejoras del programa y la decisión de continuidad o no en algunos de los municipios son recomendaciones desde los análisis realizados cuya responsabilidad está en los patrocinadores del programa.

Informe

El volumen de datos en una investigación evaluativa de enfoque cualitativo, obliga a organizar y resumir teniendo en cuenta la claridad necesaria tanto en el análisis de los datos como en la valoración del programa. Además, las diferentes audiencias requieren una precisión en su propio programa. Es decir, se requiere informes diferentes que recojan exhaustivamente los datos pero en poco espacio.

En este caso, se elaboraron 5 informes; uno global para las instituciones responsables del programa (Bienestar Social, Juventud, ambas de Diputación, y Educación) y otro para cada uno de los municipios, detallando las informaciones propias del programa en ese ámbito.

La estructura del infonne presenta en primer término las recomendaciones generales y por municipios; las conclusiones en referencia a cada objetivo de la evaluación; los resultados referidos al programa desde los objetivos y criterios/referencias, junto con la definición de la figura del educador social y la reconstrucción real del programa (lo que se hace en cada municipio) y por último, la descripción del proceso y ejemplificación de los análisis realizados.

Las ejemplificaciones no sólo se han presentado para justificar la veracidad de las conclusiones, sino que forman parte de la claridad del informe, para la comprensión del mismo por parte del lector, quedando patente en ellas la complejidad de las acciones desarrolladas a través de un programa de ayuda social.

659

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **EVALUACIÓN DE CENTROS DEL PAÍS VASCO**

RESPONSABLE DEL PROGRAMA

Departamento de Educación del Gobierno Vasco.

RESPONSABLE DE LA EVALUACIÓN

J.F. Lukas, C. Santiago, B. Munarriz, N. Moyano y M. Sedano (Dpto. MIDE. Universidad del País Vasco/Euskal Herriko Unibertsitatea).

OBJETO EVALUADO

La Red de Evaluación del País Vasco es un proyecto que trabaja de forma coordinada con un modelo similar que se está desarrollando en Madrid, Cataluña y Andalucía. Se trata de una evaluación externa de 55 centros de Secundaria de la Comunidad Autónoma del País Vasco.

OBJETIVOS DE LA EVALUACIÓN

El objetivo principal de la evaluación es dotar a los centros de secundaria de una información relevante y contextualizada acerca de su funcionamiento. De esta forma, los centros podrán conocer su situación, detectar los puntos fuertes y débiles y establecer las pautas de cambio para su mejora.

Función: Las dos funciones principales de la Red de Evaluación de Centros son las siguientes:

* Ayudar a los centros participantes a que conozcan mejor su realidad educativa y puedan de esa manera impulsar estrategias de cambio y de mejora. Por ello, aunque también tiene una vertiente de tipo sumativo, se trata sobre todo de una evaluación con carácter formativo.
* Investigar los factores responsables de la calidad de la enseñanza.

Paradigma: Los estudios sobre la eficacia de las escuelas entran dentro del campo de la investigación científica por lo que su finalidad principal es ampliar el conocimiento disponible sobre el funcionamiento de las escuelas y los factores que en él

inciden. Por otro lado los programas para mejorar la escuela tienen un objetivo más aplicado y pretenden principalmente comprobar la utilidad de determinadas estrategias de cambio. En el primer caso, aunque los resultados obtenidos son rigurosos, no se pueden trasladar directamente a la práctica. En el segundo caso, las conclusiones obtenidas son más limitadas y con menor poder de generalización, pero son aplicables en el contexto.

En los últimos años ha habido intentos de acercar estos dos paradigmas y generar una dinámica de apoyo mutuo. Este es el caso de la Red de Evaluación de Centros del País Vasco cuyos dos objetivos principales son investigar los factores de la calidad de la enseñanza y utilizar el conocimiento que se obtenga para que las escuelas que están participando en la red puedan mejorar su funcionamiento.

Modelo: Este modelo está organizado en cuatro niveles: entrada, contexto, procesos y resultados. En cada uno de ellos se establecen varias dimensiones.

El nivel de *entrada* incluye los resultados iniciales de los/as alumnos/as al incorporarse a la educación secundaria, lo que se considera fundamental para analizar los cambios que se producen a lo largo del tiempo y determinar con mayor fiabilidad los cambios que se producen a lo largo del tiempo y determinar con mayor fiabilidad el valor añadido de la escuelas.

El nivel de *contexto* se refiere al contexto sociocultural. Todos los estudios ponen de relieve la influencia del contexto sociocultural en los resultados que obtienen los/as alumnos/as, aunque su valor predictivo es menor si se han controlado también los resultados iniciales.

El nivel de *procesos* incluye dos ámbitos: los de organización y funcionamiento del centro y los más específicos del aula. Por lo que se refiere a los procesos institucionales se incluye el conocimiento del funcionamiento del centro, la participación, el clima de trabajo, la valoración del equipo directivo y del trabajo en el seminario didáctico, las relaciones con el alumnado y las expectativas sobre el centro escolar. En relación con los procesos de aula se ha tenido en cuenta especialmente la planificación de la enseñanza, la innovación en la evaluación del alumnado, al amplitud de los contenidos y la capacidad para mantener un ambiente de trabajo ordenado.

El nivel de *resultados* se centra principalmente en el alumnado, pero incluye también la opinión de los padres y madres y del profesorado. En relación al alumnado se evalúan los aprendizajes en las principales áreas curriculares (matemáticas, lengua, euskara, ciencias sociales y experimentales y tecnología), así como sus habilidades metacognitivas, estrategias de aprendizaje y actitudes. También completan un cuestionario en el que manifiestan su opinión sobre el funcionamiento del centro, las calificaciones que obtienen, la preparación recibida, sus profesores/as, tutores/as y compañeros/as. Los padres y madres expresan su opinión a través de un cuestionario sobre el funcionamiento del centro, la atención recibida, la disciplina que hay en la escuela, las calificaciones de sus hijos/as y las actividades extraescolares. La valoración del profesorado incluye la satisfacción por el funcionamiento del centro y por las condiciones en las que realiza su trabajo.

METODOLOGÍA DE LA EVALUACIÓN

Técnicas de recogida de información

Para la medición de la mayoría de los niveles y dimensiones considerados en la evaluación se utilizan cuestionarios cerrados y pruebas de ítems de elección múltiple. Todas las pruebas son contrastadas empíricamente siguiendo los procedimientos habituales de la teoría clásica de los tests y de la teoría de respuesta a los ítems, garantizando así su fiabilidad y validez.

Criterios utilizados para valorar la información

Las características de la información obtenida son las siguientes:

— *Contextualizada*. Se controla el rendimiento inicial del alumnado así como el nivel socioeconómico. La información que se proporciona a los centros tiene en cuenta siempre estos datos iniciales.
— *Comparada*. Cada escuela recibe los resultados que obtiene en todas las dimensiones estudiadas en comparación con la media de los resultados obtenidos por las escuelas de su mismo contexto sociocultural y por la totalidad de las escuelas participantes en la red. Estos datos se proporcionan tanto a nivel de comunidad como a nivel de todo el Estado.
— *Confidencial*. La información de cada centro sólo es conocida por el propio centro.
— *Objetiva*. La información que se proporciona procede de los cuestionarios y pruebas estandarizadas que se han aplicado en el centro.
— *Interpretada por los propios centros y profesores*. La información facilitada debe ser analizada, valorada e interpretada por cada centro.
— *Amplia y convergente*. La información que se obtiene no se refiere solamente a los resultados académicos del alumnado, sino que recoge también sus actitudes y estrategias de aprendizaje, los procesos de centro y aula y las valoraciones de padres y madres, profesores/as y alumnos/as. De esta manera es posible presentar los puestos de vista de los distintos sectores de la comunidad educativa que no siempre son coincidentes.
— *Continuada*. Los centros deben participar en la evaluación durante cuatro años. De esta forma hay un seguimiento longitudinal.

Referencias

Las referencias en las que se basa la evaluación para la interpretación de las puntuaciones obtenidas son de dos tipos:

* Comparación transversal: Se trata de una interpretación de tipo normativo. Cada centro recibe las medias obtenidas en cada una de los niveles y dimensiones analizados. Junto a estas medias se le ofrecen otras cuatro medias que le sirven de contraste:
  — Media de los centros del País Vasco de su mismo contexto sociocultural
  — Media de todos los centros del País Vasco
  — Media de los centros del Estado de su mismo contexto sociocultural
  — Media de todos los centros del Estado
* Comparación longitudinal: Como se ha mencionado anteriormente, los centros participantes asumen el compromiso de participar en la evaluación durante cuatro años. A partir del segundo año, además de las comparaciones transversales que se suceden durante los cuatro años, es posible comparar a cada centro consigo mismo. Es decir se analiza la evolución del centro en las diferentes dimensiones consideradas. Para ello se toman como base los estudios sobre el valor añadido.

De cara al curso 2000/2001 se ha abierto una nueva línea de investigación para establecer estándares y puntos de corte en las pruebas objetivas utilizadas que nos permitan realizar también comparaciones de tipo criterial.

Técnicas de registro y análisis de la información

La mayor parte de la información recogida es de tipo cuantitativo y por ello se realizan los análisis estadísticos adecuados para cada caso. Al cabo de los cuatro años de vigencia de la evaluación se dispondrá de los siguientes grupos para realizar las comparaciones:

Año 1º	Año 2º	Año 3º	Año 4º
1 a	2 a		4 a
2 b		4 b	
	4 c		
4 d			
	1 e	2 e	
		1 f	2 f
			1 g

Los números 1, 2, 4 se refieren al curso de los/as alumnos/as (al alumnado de 3' de ESO no se le aplican las pruebas). Los subíndices con letra hacen referencia a diferentes cohortes de alumnos/as: a, primera cohorte: b, segunda cohorte, etc. La coincidencia con el subíndice indica que se trata del mismo alumnado, permitiendo la comparación longitudinal. La coincidencia en el número permite la comparación transversal. Las comparaciones longitudinales pueden realizarse según el esquema

presentado en las filas de la tabla. Paralelamente pueden hacerse comparaciones transversales dentro de cada casilla de la tabla con los centros de los diferentes contextos. Los efectos de cohorte pueden examinarse con comparaciones entre los sujetos de 1°, 2° y 4° de los años sucesivos.

Tipos de decisiones

Los resultados obtenidos son enviados a los centros en donde el equipo directivo, la comisión de coordinación pedagógica y los seminarios didácticos, la analizan, discuten e interpretan. También los datos generales que se obtienen del análisis estadístico de todos los datos recogidos ayuda al centro a valorar los suyos propios. El valor predictivo de los resultados iniciales, la influencia del contexto sociocultural o del tamaño del centro junto con la influencia del resto de las variables estudiadas ofrece un marco más amplio y contrastado para que cada centro interprete sus datos específicos. A partir de estos análisis, los centros adoptan decisiones que traten de resolver alguno de los problemas detectados o mejorar el funcionamiento del centro. En los años siguientes, el centro puede comprobar el efecto de las iniciativas que ha adoptado a través de la valoración que cada sector de la comunidad educativa vuelve a expresar.

Informe

Cada centro recibe un informe confidencial correspondiente a los datos obtenidos en su centro en comparación con el resto de centros participantes en la red de evaluación. Se trata de un infonne completo y exhaustivo de tipo descriptivo en el que no aparecen ni valoraciones ni interpretaciones puesto que éstas deberán ser hechas por el propio centro. Este informe escrito se presenta mediante una reunión/discusión con el profesorado en el que además se recogen las sugerencias de los centros.

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 691-700

UNIVERSIDAD DE SALAMANCA

María Esperanza Herrara García
Departamento de Métodos de Investigación y Diagnóstico en Educación

PRIMER INFORME

TÍTULO DEL PROGRAMA: **MEJORAR NUESTRA COMPRENSIÓN Y EXPRESIÓN ESCRITA DE TEXTOS EXPOSITIVOS**

RESPONSABLES DEL PROGRAMA

Dña. Azucena Hernández Martín y Dña. Anunciación Quintero Gallego

RESPONSABLES DE LA EVALUACIÓN

Dña. Azucena Hernández Martín y Dña. Anunciación Quintero Gallego

OBJETO EVALUADO

Tres subprogramas dirigidos a mejorar la comprensión lectora y la producción escrita de alumnos de 3º de la E.S.O. Los programas se orientan a instruir en una serie de estrategias cognitivas y metacognitivas necesarias en los procesos de comprensión y producción escrita, por lo que se incluyen en el contexto de los programas de enseñar a pensar.

OBJETIVOS DE LA EVALUACIÓN

Diseñar, aplicar y comprobar el grado de validez de tres programas dirigidos a desarrollar y/o mejorar las estrategias necesarias para una adecuada comprensión y producción de textos.

Analizar los efectos diferenciales que la implementación de los programas puede tener en la mejora de las capacidades comprensivas y expresivas de alumnos de 3º de E.S.O. y, por consiguiente, en su rendimiento académico.

Función: formativa

Paradigma: cuantitativo

Modelo: CIPP

METODOLOGÍA DE LA EVALUACIÓN

Diseño: Diseño de grupos de control no equivalentes con medidas pretratamiento y postratamiento, pertenecientes a la metodología cuasiexperimental.

Técnicas de recogida de datos

Pruebas estandarizadas, aplicadas durante la fase pretest:
TEA 2
CETI
Pruebas específicas construidas para adaptarse a los objetivos de la investigación, aplicadas en las fases pretest y postest:
Cuatro pruebas de medición de la comprensión lectora.
Pruebas de medición de la expresión escrita.
Escala de estrategias cognitivas y actitudes hacia la expresión escrita.
Acta de calificaciones y capacidades de los alumnos, recogidas en las dos fases.

Criterios utilizados para valorar la información

Eficacia y efectividad.
Referencias: progreso sobre el propio programa.

Técnicas de registro y análisis de la información

Para el registro y análisis de la información, se empleó el paquete estadístico SPSS para Windows 95. La versión empleada fue la 7.5.

Los tipos de análisis aplicados a la información recogida fueron los siguientes:

En la fase pretest: estadística descriptiva (puntuaciones mínimas, máximas, porcentajes, media y desviación típica de cada una de las variables) y estadística inferencial (ANVA de un factor y test de Scheffé, Ji cuadrado y «t» para muestras independientes).

En la fase postest se realizaron los mismos análisis, incluyendo también un análisis de covarianza.

En la fase pretest-postest se aplicó una prueba t para muestras relacionadas y una ANVA de dos factores.

666

Tipos de decisiones

A partir de los resultados obtenidos se demostró la validez de los tres programas diseñados. Se concluyó que:

Una intervención específica orientada al desarrollo y empleo de los procesos y estrategias implicadas en la composición de textos variados, potenciaba la calidad de la expresión escrita de los alumnos

Una intervención específica orientada al desarrollo y empleo de las estrategias necesarias para construir una representación coherente del contenido del texto, contribuía a mejorar la comprensión lectora, y

Una intervención específica y conjunta en estrategias de comprensión lectora y expresión escrita mejoraba el desempeño de ambas competencias, aunque no de forma generalizada en todos los alumnos.

Sin embargo, teniendo en cuenta que una de las finalidades del trabajo se orientaba a comprobar cuál de los tres programas contribuía en mayor medida a incrementar la competencia comprensiva y expresiva de los alumnos, tomamos la decisión de decantarnos por el programa «Mejorar nuestra comprensión de textos expositivos», al haberse constituido como el más potente para mejorar ambas competencias en todos los alumnos del grupo, incluidos aquellos que manifestaron un menor rendimiento en las mismas desde el principio así como al haber demostrado que poseía mayores efectos sobre el rendimiento académico.

Informe: Tesis Doctoral

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **PENSAR PARA ESCRIBIR. UN PROGRAMA DE ENSEÑANZA PARA LA COMPOSICIÓN ESCRITA**

RESPONSABLE DEL PROGRAMA:

Mª del Pino Lecuona Naranjo

RESPONSABLE DE LA EVALUACIÓN:

Directora: Mª del Pino Lecuona Naranjo

Investigadores: Mª Jesús Miranda Velasco, Mª José Rodríguez Conde, Mª Cruz Sánchez Gómez, Pilar Sarto Martín, Jesús Valverde Berrocoso

OBJETO EVALUADO:

Programa de enseñar a pensar

OBJETIVOS:

La necesidad de trabajar en estudios que favorezcan la elaboración, aplicación y valoración de programas específicos de enseñanza de la composición escrita, en los que se tengan en cuenta los procesos cognitivos que subyacen a su ejecución, constituye la finalidad fundamental de este trabajo. De ahí que nuestro objetivo general sea el diseño de estrategias de enseñanza que favorezca la mejora de la calidad de la expresión escrita del alumnado en las modalidades textuales descriptiva, narrativa y expositiva. Este objetivo puede desglosarse en las siguientes fases:

1ª. Descubrir cuáles son las estructuras subyacentes en los textos elaborados por los niños y las niñas de 6º de Educación Primaria.
2ª. Analizar la composición escrita de los/las escolares de este último curso de Educación Primaria en colegios públicos céntricos y periféricos.
3ª. Analizar y sistematizar los errores más frecuentes en la elaboración de textos escritos.
4ª. Diseñar un programa de intervención a partir de modelos cognitivos, lingüísticos e interactivos o sociales.
5ª. Aplicar el nuevo programa en colegios de la zona periférica y de la zona centro de la ciudad, controlando el rendimiento de los alumnos en la materia de lenguaje.
6ª. Contrastar los resultados entre una metodología tradicional de aprendizaje de la composición escrita y el nuevo diseño elaborado para el alumnado de 6º de Educación Primaria.
7ª. Establecer algunas conclusiones acerca de la enseñanza de la expresión escrita que orienten al profesorado para la mejora de su actuación docente.

Función: Formativa

Paradigma: Cuantitativo (investigación cuasiexperimental)

Modelo: CIPP

METODOLOGÍA DE LA EVALUACIÓN:

Diseño: diseños de grupos de control no equivalentes con medidas pretratamiento y postratamiento).

Técnicas de recogida de información:
— Cuestionario inicial
— TEA y Nota profesor de Lenguaje
— Factor G
— Lista de control de estrategias. metacognitivas
— Pautas de evaluación de composición.escrita

Criterios utilizados para valorar la información:

Sobre el grado de validez interna y externa de este diseño, aspecto que nos ha preocupado desde el inicio de la investigación, se debe señalar lo siguiente:

Por un lado, en lo que se refiere a la validez interna, a la capacidad que el diseño establecido ha tenido sobre el control del error experimental, opinamos que, aún dentro de las limitaciones de este tipo de estudio, los instrumentos de medida utilizados han sido suficientemente sensibles, válidos y fiables, como para detectar los cambios efectuados y las diferencias entre grupos en aquellos aspectos que nos ha interesado medir.

Por otro lado, al plantearnos las cuestiones relativas a la validez externa, es decir, al poder de generalización de estos resultados, procedentes de alumnos y alumnas de dos centros concretos, al resto de discentes que conformarían la población, hemos comprobado, tras los análisis inferenciales efectuados, que sí aparecen indicios que permiten su transferencia.

Referencias: progreso sobre el propio programa

Técnicas análisis de la información:

La preparación cuidadosa de los datos de la variable dependiente, composición escrita, se realizó a partir de la puntuación otorgada por un sistema de jueces. Se han realizado análisis estadísticos descriptivos, correlacionales e inferenciales, en respuesta al diseño de investigación utilizado. El contraste de hipótesis se ha resuelto a partir de pruebas de «t» de diferencia de medias para muestras independientes y correlacionadas y ANCOVA, en el caso de controlar el efecto de variables correlacionadas con la variable dependiente.

Tipos de decisiones:

Una metodología de enseñanza basada en la utilización de estrategias cognitivas y lingüísticas incrementa el nivel de calidad de la expresión escrita del alumnado.

La enseñanza tradicional de la composición escrita produce niveles más bajos de rendimiento en este alumnado, que si la comparamos con la metodología de investigación empleada en este estudio.

La intervención, efectuada mediante la presentación a través de (1) motivación, (2) modelado y (3) entrenamiento en tareas metacognitivas, así como el material didáctico

elaborado ad hoc, parece que influye en una mejora en la producción escrita de los alumnos en sexto curso de Educación Primaria (11-12 años). La valoración efectuada después del entrenamiento en cada uno de los textos y, la realizada un tiempo después, parece que así lo corrobora.

La mejora, con este tipo de intervención, se aprecia de manera más ostensible en aquellos alumnos y alumnas de capacidad cognitiva media y alta, en comparación con los/las estudiantes de bajo nivel. Las características como mayor interés por la lectura, situación socioeconómica favorable, etc., pueden propiciar un mejor aprovechamiento de este tipo de intervención. Creemos que para alumnos de nivel medio y alto los programas intensivos pueden ser adecuados; en cambio, en escolares menos favorecidos, la intervención debería ser más prolongada en el tiempo, volviendo a insistir en aspectos que no llegan a comprender desde el primer momento.

Los procesos cognitivos implicados en la elaboración de textos expositivos indican la necesidad de su entrenamiento en edades más avanzadas que las consideradas en este estudio. Sería conveniente introducir dicha modalidad en cursos posteriores, pues consideramos que es conveniente retrasarlo hasta el primer ciclo de la Etapa de Educación Secundaria. Sin embargo, se debería reforzar en la etapa de Educación Primaria la lectura y la escritura de textos de estructura narrativa, puesto que los chicos y las chicas se decantan y prefieren leer relatos preferentemente de aventuras.

Como este tipo de diseño de intervención parece que está exigiendo unos condicionantes sobre conocimientos previos, y como el programa sólo hace incidencia en aspectos de coherencia y cohesión textual, sugerimos: (a) trabajar previamente los aspectos de vocabulario, tiempos verbales, ortografía, etc., y/o (b) continuar desarrollando y ampliando este programa a fin de integrarlos en el mismo.

Otra vía de trabajo se centraría en la utilización de nuevas tecnologías tales como el procesador de textos y la telemática valorando su incidencia y utilidad en los procesos de enseñanza-aprendizaje de la expresión escrita.

Para concluir, habiendo comprobado la mejora del alumnado en las habilidades de escritura tras la implementación del programa, animamos a todos los interesados en el tema a aplicarlo, a fin de contrastar nuestros resultados con los que pudieran obtenerse en otros contextos.

Informes:

Esta investigación fue publicada por el Centro de Investigación y Documentación Educativa en 1999. De ella han derivado diferentes Cursos, Ponencias y Artículos destinados fundamentalmente a Profesorado de Primaria.

TERCER INFORME

TÍTULO DEL PROGRAMA: **PENSAR PARA ESCRIBIR. UN PROGRAMA DE ENSEÑANZA PARA LA COMPOSICIÓN ESCRITA**

RESPONSABLE DEL PROGRAMA

Mª del Pino Lecuona Naranjo

RESPONSABLE DE LA EVALUACIÓN

Directora: Mª del Pino Lecuona Naranjo
Investigadores: Mª Jesús Miranda Velasco, Mª José Rodríguez Conde, Mª Cruz Sánchez Gómez, Pilar Sarto Martín, Jesús Valverde Berrocoso

OBJETO EVALUADO

Programa de enseñar a pensar

OBJETIVOS

La necesidad de trabajar en estudios que favorezcan la elaboración, aplicación y valoración de programas específicos de enseñanza de la composición escrita, en los que se tengan en cuenta los procesos cognitivos que subyacen a su ejecución, constituye la finalidad fundamental de este trabajo. De ahí que nuestro objetivo general sea el diseño de estrategias de enseñanza que favorezca la mejora de la calidad de la expresión escrita del alumnado en las modalidades textuales descriptiva, narrativa y expositiva. Este objetivo puede desglosarse en las siguientes fases:

1ª. Descubrir cuáles son las estructuras subyacentes en los textos elaborados por los niños y las niñas de 6º de Educación Primaria.
2ª. Analizar la composición escrita de los/las escolares de este último curso de Educación Primaria en colegios públicos céntricos y periféricos.
3ª. Analizar y sistematizar los errores más frecuentes en la elaboración de textos escritos.
4ª. Diseñar un programa de intervención a partir de modelos cognitivos, lingüísticos e interactivos o sociales.
5ª. Aplicar el nuevo programa en colegios de la zona periférica y de la zona centro de la ciudad, controlando el rendimiento de los alumnos en la materia de lenguaje.
6ª. Contrastar los resultados entre una metodología tradicional de aprendizaje de la composición escrita y el nuevo diseño elaborado para el alumnado de 6º de Educación Primaria.

7ª. Establecer algunas conclusiones acerca de la enseñanza de la expresión escrita que orienten al profesorado para la mejora de su actuación docente.

Función: Formativa

Paradigma: Cuantitativo (investigación cuasiexperimental)

Modelo: CIPP

MÉTODOLOGÍA DE LA EVALUACIÓN

Diseño: diseños de grupos de control no equivalentes con medidas pretratamiento y postratamiento).

Técnicas de recogida de información:

— Cuestionario inicial
— TEA y calificación asignada por el profesor de Lenguaje
— Factor G
— Lista de control de estrategias metacognitivas
— Pautas de evaluación de composición escrita

Criterios utilizados para valorar la información

Sobre el grado de validez interna y externa de este diseño, aspecto que nos ha preocupado desde el inicio de la investigación, se debe señalar lo siguiente:

Por un lado, en lo que se refiere a la validez interna, a la capacidad que el diseño establecido ha tenido sobre el control del error experimental, opinamos que, aún dentro de las limitaciones de este tipo de estudio, los instrumentos de medida utilizados han sido suficientemente sensibles, válidos y fiables, como para detectar los cambios efectuados y las diferencias entre grupos en aquellos aspectos que nos ha interesado medir.

Por otro lado, al plantearnos las cuestiones relativas a la validez externa, es decir, al poder de generalización de estos resultados, procedentes de alumnos y alumnas de dos centros concretos, al resto de discentes que conformarían la población, hemos comprobado, tras los análisis inferenciales efectuados, que sí aparecen indicios que permiten su transferencia.

Referencias: progreso sobre el propio programa

Técnicas de análisis de la información

La preparación cuidadosa de los datos de la variable dependiente, composición escrita, se realizó a partir de la puntuación otorgada por un sistema de jueces. Se han

realizado análisis estadísticos descriptivos, correlacionales e inferenciales, en respuesta al diseño de investigación utilizado. El contraste de hipótesis se ha resuelto a partir de pruebas de «t» de diferencia de medias para muestras independientes y correlacionadas y ANCOVA, en el caso de controlar el efecto de variables correlacionadas con la variable dependiente.

Tipos de decisiones

Una metodología de enseñanza basada en la utilización de estrategias cognitivas y lingüísticas incrementa el nivel de calidad de la expresión escrita del alumnado.

La enseñanza tradicional de la composición escrita produce niveles más bajos de rendimiento en este alumnado que si la comparamos con la metodología de investigación empleada en este estudio.

La intervención, efectuada mediante la presentación a través de (1) motivación, (2) modelado y (3) entrenamiento en tareas metacognitivas, así como el material didáctico elaborado *ad hoc*, parece que influye en una mejora en la producción escrita de los alumnos en sexto curso de Educación Primaria (11-12 años). La valoración efectuada después del entrenamiento en cada uno de los textos y, la realizada un tiempo después, parece que así lo corrobora.

La mejora, con este tipo de intervención, se aprecia de manera más ostensible en aquellos alumnos y alumnas de capacidad cognitiva media y alta, en comparación con los/las estudiantes de bajo nivel. Las características como mayor interés por la lectura, situación socioeconómica favorable, etc., pueden propiciar un mejor aprovechamiento de este tipo de intervención. Creemos que, para alumnos de nivel medio y alto, los programas intensivos pueden ser adecuados; en cambio, en escolares menos favorecidos, la intervención debería ser más prolongada en el tiempo, volviendo a insistir en aspectos que no llegan a comprender desde el primer momento.

Los procesos cognitivos implicados en la elaboración de textos expositivos indican la necesidad de su entrenamiento en edades más avanzadas que las consideradas en este estudio. Sería conveniente introducir dicha modalidad en cursos posteriores, pues consideramos que es conveniente retrasarlo hasta el primer ciclo de la Etapa de Educación Secundaria. Sin embargo, se debería reforzar en la etapa de Educación Primaria la lectura y la escritura de textos de estructura narrativa, puesto que los chicos y las chicas se decantan y prefieren leer relatos preferentemente de aventuras.

Como este tipo de diseño de intervención parece que está exigiendo unos condicionantes sobre conocimientos previos, y como el programa sólo hace incidencia en aspectos de coherencia y cohesión textual, sugerimos: (a) trabajar previamente los aspectos de vocabulario, tiempos verbales, ortografía, etc., y/o (b) continuar desarrollando y ampliando este programa a fin de integrarlos en el mismo.

Otra vía de trabajo se centraría en la utilización de nuevas tecnologías tales como el procesador de textos y la telemática valorando su incidencia y utilidad en los procesos de enseñanza-aprendizaje de la expresión escrita.

Para concluir, habiendo comprobado la mejora del alumnado en las habilidades de escritura tras la implementación del programa, animamos a todos los interesados en el

tema a aplicarlo, a fin de contrastar nuestros resultados con los que pudieran obtenerse en otros contextos.

Informes

Esta investigación fue publicada por el Centro de Investigación y Documentación Educativa en 1999. De ella han derivado diferentes Cursos, Ponencias y Artículos destinados fundamentalmente a Profesorado de Primaria.

674

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 701-706

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

Ana Porto Castro
Área de Métodos de Investigación y Diagnóstico en Educación

PRIMER INFORME

TÍTULO DEL PROGRAMA: **EFECTOS DE UN PROGRAMA EN EL ÁREA DE MATEMÁTICAS CON ALUMNOS MUY CAPACITADOS.**

RESPONSABLE DEL PROGRAMA

Ana Mª Porto Castro
Área MIDE. Facultad de Ciencias de la Educación
Director: Dr. José Cajide Val

RESPONSABLE DE LA EVALUACIÓN

Ana Mª Porto Castro

OBJETO EVALUADO

Un programa en el área de matemáticas pensado especialmente para alumnos muy capacitados. El programa se centra en el desarrollo de capacidades, destrezas y habilidades del alumno. Se pretende el desarrollo y estimulación del alumno para resolver problemas y para utilizar las estrategias para su resolución.

OBJETIVOS

— Comprobar los efectos del programa y si estos son especialmente favorables para los alumnos identificados como muy capacitados.

— Comprobar si las variables consideradas para seleccionar a los alumnos como muy capacitados son predictores efectivos de los efectos del programa.

Función: Sumativa

Paradigma: Cuantitativo

Modelo: Resultados

METODOLOGÍA DE LA EVALUACIÓN

- Diseño: Se utiliza un diseño de 4 grupos (uno experimental y otro control seleccionados al azar; uno experimental y otro control de sujetos identificados como muy capacitados), pretes-postest, con una variable independiente.
- Técnicas de recogida de información: Cuestionario.
- Criterios utilizados para valorar la información: Suficiencia, evaluabilidad y eficacia del programa.
- Referencias: Progreso sobre el propio programa.
- Técnicas de registro y análisis de la información: La información se analiza de modo cuantitativo, a través del cálculo de estadísticos descriptivos y de la aplicación de las siguientes pruebas estadísticas: t de Student, Análisis de Varianza, Análisis de Regresión y Análisis de Covarianza.
- Tipo de decisiones:
- Informe: Se trata de un trabajo de Tesis Doctoral.

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **LA PRENSA DIARIA COMO RECURSO MOTIVADOR DE LA LECTURA EN LAS AULAS**

RESPONSABLE DEL PROGRAMA

Cristina Abeal Pereira.
Área MIDE. Facultad de Ciencias de la Educación.
Director: Dr. José Cajide Val.

RESPONSABLE DE LA EVALUACIÓN

Cristina Abeal Pereira.

OBJETO EVALUADO

Un programa para el aprendizaje lector en niños desmotivados, usando como recurso didáctico la prensa escrita.

OBJETIVOS

— Demostrar que un adecuado uso selectivo del periódico en el aula es una buena estrategia de incentivación para el desarrollo de la comprensión lectora.
— Descubrir diferencias relevantes entre los elementos de motivación propiciados por el uso selectivo de la prensa y por otros textos, en orden a la comprensión lectora.
— Verificar si existen diferencias entre los rendimientos en comprensión lectora de aquellos alumnos con dificultades que utilicen como recurso de motivación la prensa y el de aquellos que usan recursos más convencionales.
— Comprobar si existen diferencias entre el nivel de satisfacción lectora de los alumnos que utilizan la prensa y aquellos que utilizan un recurso más convencional.

Función: Formativa y sumativa.

Paradigma: Cuantitivo.

Modelo: Resultados.

METODOLOGÍA DE LA EVALUACIÓN

- Diseño: Se utiliza un diseño 4 grupos (un grupo experimental y otro control de alumnos con dificultades y desmotivación lectora; un grupo experimental y un grupo control seleccionados al azar, de alumnos sin dificultades ni desmotivación lectora), pretest-postest con una variable independiente
- Técnicas de recogida de información: Cuestionario y entrevista.
- Criterios utilizados para valorar la información: Viabilidad, eficacia y eficiencia.
- Referencias: Progreso sobre el propio programa y el punto de partida.
- Técnicas de registro y análisis de la información: Se calculan estadísticos descriptivos y se realiza la prueba t de Student.
- Tipo de decisiones:
- Informe: Trabajo de Tesis Doctoral.

TERCER INFORME INFORME

TÍTULO DEL PROGRAMA: **EFECTOS DEL USO DE LOS MEDIOS INFORMÁTICOS EN EL APRENDIZAJE DE LOS ALUMNOS/AS DE SEGUNDO DE LA ESO EN EL ÁREA DE TECNOLOGÍA: UN ESTUDIO DE CASO**

RESPONSABLE DEL PROGRAMA

Elisa Teresa Zamora Rodríguez.
Área MIDE. Facultad de Ciencias de la Educación.
Director: Dr. José Cajide Val.

RESPONSABLE DE LA EVALUACIÓN

Elisa Teresa Zamora Rodríguez

OBJETO EVALUADO

Se evalúa un proyecto didáctico dirigido a alumnos de 2º de ESO en el área de Tecnología, articulado en torno a un eje central consistente en la confección de un Tangram. Para la confección del Tangram es requisito previo conocer cómo es realmente; este conocimiento constituye un paso necesario, una meta intermedia en su confección.

OBJETIVOS

— Estudiar y analizar los efectos del uso de los medios informáticos en el proceso de enseñanza-aprendizaje del área de tecnología, teniendo en cuenta los objetivos, contenidos, estrategias metodológicas, tareas y actividades a desenvolver en esta área.

Función: Formativa

Paradigma: Cualitativo

Modelo:

METODOLOGÍA DE LA EVALUACIÓN

- Diseño: Estudio de caso único
- Técnicas de recogida de información: Cuestionario, Entrevista, Observación y documentos escritos
- Criterios utilizados para valorar la información: Evaluabilidad, Viabilidad y Eficacia

- Referencias: Progreso sobre el programa y consecución de los objetivos del área
- Técnicas de registro y análisis de la información: Triangulación de técnicas y métodos, cálculo de fiabilidad del cuestionario, análisis del desarrollo instructivo de los alumnos en el área de tecnología, y análisis descriptivo.
- Informe: Trabajo de Tesis Doctoral.

CUARTO INFORME

TÍTULO DEL PROGRAMA: **PROGRAMA DE APROVECHAMIENTO INTEGRAL DEL ESPACIO ESCOLAR Y DE LA MEJORA DE LA CALIDAD DE LA ENSEÑANZA**

RESPONSABLE DEL PROGRAMA

Neuza Celina Canabarro Elizeire
Área MIDE, Facultad de Ciencias de la Educación
Director: Dr. José Cajide Val

RESPONSABLE DE LA EVALUACIÓN

Neuza Celina Canabarro Elizeire

OBJETO EVALUADO

Programa implantado en escuelas públicas del Estado Rio do Sul desde 1992-1994, diseñado para eliminar la falta de plazas en las escuelas, la actualización y el perfeccionamiento de los recursos humanos de la red pública estatal y para atender la demanda escolar de aproximadamente 200.000 excedentes.

OBJETIVOS

— Comprobar si el programa permite mejorar las condiciones que resuelvan la falta de plazas en la red pública estatal.

Función: Sociopolítica

Paradigma: Cuantitativo.

Modelo: Proceso-Producto.

METODOLOGÍA DE LA EVALUACIÓN:

- Diseño
- Técnicas de recogida de información: Cuestionario por correo y entrevista.
- Criterios utilizados para valorar la información: Viabilidad y efectividad
- Referencias: Superioridad de esta alternativa sobre otras adoptadas en el Estado y en los municipios, en las últimas décadas.
- Técnicas de registro y análisis de la información: Estadísticos descriptivos y Análisis de varianza
- Informe: Trabajo de Tesis Doctoral

680

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 707-712

UNIVERSIDAD DE SEVILLA

Antonio Matas Terron
Área de Métodos de Investigación y Diagnóstico en Educación

En la comunidad andaluza existe el Plan Andaluz de Investigación (P.A.I.). En él se integran y reconocen un gran número de grupos de investigación que abarcan todos los ámbitos del conocimiento así del desarrollo científico y tecnológico. A través de este plan, los miembros del área MIDE encuentran la posibilidad de desarrollar ciertos proyectos de investigación con cierto reconocimiento «oficial», a través de la creación de grupos de investigación. Los grupos suelen ser interdisciplinares, integrados por participantes de uno o más centros, y procedentes tanto del ámbito universitario como de centros educativos de secundaria, primaria, etc., empresas privadas y públicas, administración pública, etc.

Dentro de este marco general, se desarrollan los proyectos de investigación en el área MIDE de la Universidad de Sevilla. Respecto al objeto de esta exposición, actualmente se encuentran en marcha dos proyectos de investigación en los que participan miembros del área. Por una parte, el proyecto, denominado «Atando Cabos». Realizado por el grupo de investigación MIDO (Métodos de Investigación y Diagnóstico en Orientación). Su director es el Doctor Don Víctor Álvarez Rojo, catedrático del área MIDE. El equipo principal está formado por: Dr. Don Eduardo García, Dra. Da. Soledad Romero, y Dr. Don Javier Gil. Se trata de un programa informatizado de auto análisis de los métodos de trabajo para alumnos universitarios. Es un programa de desarrollo educativo de corte integrado en cuanto al paradigma de referencia. En general sigue las líneas esenciales del modelo Stufflebeam de evaluación.

Un segundo programa es «Imagen y Fantasía». Llevado a cabo dentro del Grupo de Investigación Suroeste de la Universidad de Huelva y donde hay una participa-

ción significativa de componentes pertenecientes a la Universidad de Sevilla. La Dra. Jerónima Ipland García dirige el grupo de investigación, siendo componentes principales de la línea de investigación metacognitiva el Dr. Narciso Barrero González y el Dr. Miguel Mª Reyes Rebollo. Se trata de un programa metacognitivo de estrategias de comprensión lectora para ciclo medio de educación primaria para alumnos normoyentes e hipoacúsicos en régimen de integración. El programa integra los paradigmas cuantitativo y ecológico, siguiendo un modelo metacognitivo.

A continuación se expone una ficha resumen de ambos proyectos, donde se han indicado las características principales de los mismos.

PRIMER INFORME

TÍTULO DEL PROGRAMA: **«ATANDO CABOS»: SISTEMA INFORMATIZADO DE AUTOANÁLISIS DE LOS MÉTODOS DE TRABAJO PARA ALUMNOS UNIVERSITARIOS**

RESPONSABLE DEL PROGRAMA

ENTIDAD: Grupo MIDO
DIRECTOR: Víctor Álvarez Rojo

RESPONSABLE DE LA EVALUACLÓN

DIRECTOR: Víctor Álvarez Rojo
EQUIPO PRINCIPAL: Eduardo García Jiménez, Javier Gil Flores y Soledad Romero Rodríguez.

Objeto evaluado: OTROS: Programa de desarrollo educativo.

Paradigma: No puede hablarse de una adscripción clara a ninguno.

Modelo: No se sigue ningún modelo determinado. No obstante, en buena medida el concepto y el desarrollo de la evaluación se apoyan en las ideas de Stufflebeam, de modo que la evaluación nos ha orientado en la toma de decisiones sobre el programa (contenido, diseño, usuarios, etc.).

METODOLOGÍA

Diseño:

El modelo comparativo previo a la valoración del programa se ha apoyado en un estudio descriptivo y en otro de tipo ex-post-facto.

Técnicas de recogida de información

Cuestionario y escala de opinión, Diferencial Semántico y Observación (a partir de un sistema de categorías semicerrado).

Criterios de valoración de la información

Relevancia de los contenidos, claridad, coherencia, capacidad de motivación, adecuación a los usuarios, versatilidad, viabilidad, eficacia, eficiencia., facilidad de uso, calidad del entorno audiovisual, satisfacción del usuario y evaluabilidad.

Técnicas de registro y análisis de la información

Cédula y hoja de registro. Estadísticos descriptivos y de contraste paramétrico y no paramétrico. Análisis factorial exploratorio (estudio de la estructura de la Escala y el Diferencial Semántico).

Tipos de decisiones

Modificaciones relacionadas con el contenido del programa y con los interfaces usuario-ordenador.

Informe de evaluaciones

DESTINATARIOS: Diseñadores del programa y C.I.D.E.

ESTRUCTURA: Formaba parte de la Memoria Final de Investigación. En dicho informe se especifican: los objetivos de la evaluación, los términos de comparación, el diseño, el desarrollo de la fase de aplicación, la muestra participante, los procedimientos de recogida y análisis de datos, los resultados, las valoraciones realizadas y las decisiones adoptadas en relación con el programa.

AMPLITUD: 49 páginas.

Informe

El proceso de evaluación aquí resumido nos ha permito concluir con una valoración positiva del programa «ATANDO CABOS», tanto en los aspectos relativos a su diseño y aplicación como a los resultados que permite alcanzar a los alumnos universitarios matriculados en el primer curso, a quienes está dirigido dicho programa.

En relación con el diseño, hemos podido comprobar que los contenidos se adaptan a las necesidades de los usuarios y son adecuados para ellos, así mismo las estrategias utilizadas en el programa resultan válidas en un entorno informatizado y podemos decir que el propio diseño del *software* facilita la interacción con el usuario.

Respecto a la aplicación del programa entendemos que resulta viable en centros universitarios de las diferentes titulaciones; es fácil de usar por los estudiantes que sienten a gusto trabajando con esta aplicación informática.

Los resultados obtenidos inmediatamente después de la aplicación del programa nos muestran reacciones de los estudiantes que indican que «ATANDO CABOS» propicia el aprendizaje, resulta útil para el alumno, arroja un balance positivo en términos de coste-beneficio y, en fin, resulta satisfactorio para sus usuarios.

Como consecuencia de la evaluación se detectaron algunas deficiencias en la versión beta del programa (la utilizada en la aplicación a los alumnos) que han sido corregidas. Dichas deficiencias afectaron sobre todo al contenido y a los interfaces hombre-máquina.

SEGUNDO INFORME

TÍTULO DEL PROGRAMA: **«IMAGEN Y FANTASÍA»**

RESPONSABLES DEL PROGRAMA

ENTIDAD: Grupo de Investigación Suroeste de la Universidad de Huelva

DIRECTORES/AUTORES: Dr. Narciso Barrero González, Dra. Jerónima Ipland García y Dr. Miguel Mª Reyes Rebollo

RESPONSABLES DE LA EVALUACIÓN

Los autores

Objeto evaluado: Programa metacognitivo de estrategias de comprensión dependientes de la tarea lectora para ciclo medio de educación primaria con alumnos normoyentes e hipoacúsicos en régimen de integración.

Paradigma: Mediacional, centrado en el alumno con integración del paradigma cuantitativo y el ecológico.

Modelo: Orientación metacognitiva dependiente de la tarea escolar.

METODOLOGÍA DE LA EVALUACIÓN

Diseño

Una fase pretest, creando la línea base y una fase postest, estableciendo los resultados finales de una acción diferencial

Técnicas de recogida de información

Pretest-postest N: 108 (Factor «g» de Catell, 1992; prueba de personalidad de Gillis, 1989; comprensión lectora «Cloze», Suárez y Meara, 1985 y prueba de procesos cognitivos estratégicos de comprensión lectora, de Reyes y Barrero, 2000).

Criterios de valoración de la información

Aumento significativo del repertorio de estrategias metacognitivas dependientes de tareas e incremento cognitivo independiente de tareas

Técnicas de registro y análisis de la información

Diseño de dos grupos con una variable independiente, el Programa «Imagen y Fantasía», y con seis variables dependientes: inteligencia general, factor verbal, vocabulario, comprensión lectora, procedimientos metacognitivos y conocimiento de estrategias metacognitivas de comprensión lectora

«t» de Student con dos muestras relacionadas para contraste de hipótesis nula en la que la muestra procede de una población de medias X e Y iguales.

Informe (destinatarios, estructura y amplitud)

20 Unidades que se desarrollan a lo largo de 19 semanas más el Test metacognitivo TSD-20. El hilo conductor es su protagonista «Tonchlto», un niño de la misma edad que los alumnos que los acompaña por 19 centros de interés (aventuras) para el autocontrol ejecutivo.

Resultados de la evaluación efectuada

TERCER CURSO
Variable Comprensión Lectora (SPSS: t-value: 13,29 2-tail sig p 0,000)
Variable de Estrategias Metacognitivas (SPSS: t-value: 14,16 2-tail sig p 0,000).
Variable de Inteligencia General: (SPSS: t-value: 12,56 2-tail sig p 0,000)
Variable de Procedimientos Metacognitivos: (SPSS: t-value: 13,79 2-tail sig p 0,000)
Variable de Vocabulario (SPSS: t-value: 10,30 2-tail p 0,000)
Variable Factor Verbal: (SPSS: t-value:20,15 2-tail p 0,000)

CUARTO CURSO:
Variable Comprensión Lectora: (SPSS: t-value: 10,34 2-tail p > 0,000)
Variable de Estrategias Metacognitivas: (SPSS: t-value: t-value: 10,71 2-tail p 0,000)
Variable Inteligencia: (SPSS: t-value: 16,21 2-tail, p 0,000)
Variable de Procedimientos Metacognitivos: (SPSS: t-value: 19,54 2-tail p 0,000)
Variable de Vocabulario: (SPSS: t-value: 19,54 2-tail p 0,000)
Variable Factor Verbal: (SPSS: t-value: 21,75 2-tail p 0,000)

CONCLUSIONES

Como puede comprobarse en los datos estadísticos, las mejoras son significativas en las seis variables dependientes enunciadas con lo que nuestra investigación corrobora estudios (Bakaer y Brown, 1984; Winograd, 1982; Repetto, 1990) los cuales habían establecido previamente tanto fuera de nuestras fronteras como en el área de influencia del castellano que son posibles estas mejoras de la conducta estratégica, mediante programas metacognitivos de diseño mediacional. Las mejoras en los contenidos metacognitivos implican que las estrategias se han asimilado de forma ejecutiva pasando a ser procesos cognitivos auto-dirigidos integrados en un repertorio personal de estrategias de resolución de problemas de comprensión.

686

Revista de Investigación Educativa, 2000, Vol. 18, n.º 2, págs. 713-721

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

Catalina Martínez Mediano
Departamento de Métodos de Investigación y Diagnóstico en Educación

AÑO

Las fechas de terminación de las siete investigaciones que recogemos van desde 1991 a 1998. La 1) y 2) de 1991, la 3) de 1993, la 4) de 1994, y la 5), 6) y 7) de 1998.

TÍTULOS DE LOS PROGRAMAS

1) Factores de eficacia de los Centros Asociados de la UNED.
2) Proyectos experimentales para la Reforma de la Enseñanza de la Educación Infantil
3) Educación vial para educadores extraescolares.
4) Evaluación formativa de los alumnos de Pedagogía Experimental (curso de adaptación) de la UNED, a través del videotexto interactivo.
5) Desarrollo de habilidades superiores de pensamiento: Análisis, Síntesis y Evaluación.
6) Necesidades y servicios de orientación universitaria en la Comunidad de Madrid.
7) Proyectos educativos desarrollados en contextos escolares multiculturales.

RESPONSABLE DEL PROGRAMA

En la mayoría de los casos los programas pertenecen a instituciones —UNED (1, 3 y 6), MEC (2 y 7), Dirección General de Tráfico, Comunidad de Madrid— o a los propios investigadores (4 y 5).

RESPONSABLE DE LA EVALUACIÓN

1) C. Martínez Mediano, 2) T. Aguado Odina, 3) R. Pérez Juste (Director), E. De Lara y A. Sebastián, 4) R. Pérez Juste (Director), J.L. García Llamas, M.A. González Galán, C. Martínez Mediano, E. De Lara y N. Ruipérez Losada., 5) B. Álvarez González, 6) M.F. Sánchez García, 7) T. Aguado Odina.

OBJETO EVALUADO

En 5 de las 7 investigaciones son programas en funcionamiento de las instituciones. El 4) y 5) son programas elaborados por los investigadores. Estos son los objetos de evaluación:

1) Los Centros Asociados de la UNED.
2) Proyectos experimentales y no experimentales en el marco del Plan Experimental para la Reforma de la Enseñanza Infantil en España.
3) Programa de Educación Vial para educadores extraescolares.
4) El videotexto interactivo como herramienta para la dirección y evaluación formativa del aprendizaje, y mejora de actitudes del estudiante hacia la UNED.
5) Programa cognitivo diseñado como recurso de enriquecimiento para alumnos de altas capacidades de cinco a siete años.
6) Servicios de orientación de las universidades de la Comunidad de Madrid en relación con las necesidades de orientación del alumnado. Análisis de sus estructuras organizativas, objetivos, funciones y actividades.
7) Proyectos y prácticas educativas desarrollados en centros públicos de enseñanza obligatoria con alumnado diverso culturalmente.

OBJETIVOS

Los objetivos de las investigaciones evaluativas se centran en variables de entrada, contexto, proceso y resultados, con especial incidencia en los aspectos organizativos y procesuales de los programas, con la finalidad de explicar sus características y valorar su calidad, haciendo propuestas de mejora. En concreto, las investigaciones evaluativas que comentamos, se fijaron los siguientes objetivos:

1) Analizar las variables de los Centros Asociados de la UNED que pueden incidir en el rendimiento de los alumnos, como son la estructura física y funcional del centro, sus recursos económicos, organizativos y personales, los programas de estudio, las de los alumnos, los procesos de enseñanza-aprendizaje, y el clima socioeducativo y el contexto UNED.
2) Análisis comparativo de diferentes proyectos de educación infantil desarrollados en el marco del Plan Experimental para la Reforma de la Educación Infantil, y proyectos no pertenecientes a dicho Plan en orden a identificar las variables de entrada y proceso que diferencian entre programas reforma/no reforma y su

incidencia en los resultados de los alumnos; análisis de los efectos diferenciales de los programas cuando se adoptan como criterios los objetivos generales establecidos en la Reforma de la Educación Infantil (MEC, 1986/1990).

3) Valorar el programa de Educación Vial para educadores extraescolares en cuanto al grado en que responde a las demandas sociales, su calidad técnica y su adecuación para el aprendizaje a distancia y su utilidad para el ejercicio profesional. Valorar el desarrollo del programa en cuanto a la atención y el apoyo ofrecido /recibido, su adecuación a la planificación, la calidad de las actividades y el valor pedagógico de la evaluación. Valorar su eficacia.
4) Ampliar los elementos de relación entre el estudiante y profesor de la UNED. Mejorar el rendimiento académico mediante la dirección del aprendizaje y su seguimiento a través del videotexto interactivo. Mejorar la satisfacción de los alumnos con su proyecto educativo y la institución.
5) Desarrollo de las habilidades de pensamiento de análisis, síntesis y evaluación en alumnos de altas capacidades. Asimilación y transferencia de las estrategias aprendidas. Desarrollo en los alumnos de la capacidad de autoevaluación. Desarrollo de un recurso de enriquecimiento para el primer ciclo de Primaria.
6) Identificar y valorar las necesidades de orientación de los universitarios madrileños, en función de variables como edad, sexo, situación sociolaboral, estudios, universidad, circunstancias de elección de estudios, satisfacción con los estudios y expectativas profesionales. Análisis comparativo de los servicios/sistemas de orientación de las universidades objeto del estudio. Análisis de los ajustes y desajustes entre las necesidades detectadas y las intervenciones de los servicios de orientación.
7) Analizar y valorar las prácticas educativas desarrolladas en contextos escolares multiculturales con objeto de identificar modelos educativos y pautas de actuación desarrollados y valorar la efectividad de los mismos en función de objetivos derivados de modelos educativos interculturales. Formular un modelo que describa las relaciones entre las variables estudiadas. Elaborar propuestas de actuación coherentes con el enfoque intercultural propuesto.

FUNCIÓN

La función evaluativa que predomina es la formativa. Todos los investigadores evaluativos investigan con la finalidad de conocer mejor el objeto evaluado para poder realizar propuestas de mejora argumentadas. Junto a esta finalidad formativa, la función de informar y rendir cuentas a la sociedad próxima a los programas es, asimismo, realizada, siendo esto una característica consustancial con la finalidad de mejora de los programas y de las personas implicadas en los mismos.

PARADIGMA

En la totalidad de los trabajos que comentamos predomina un enfoque paradigmático mixto, ecléctico, o de complementariedad metodológica. En cierto modo se trata

de una «cualitativización de las investigaciones cuantitativas». Los investigadores del ámbito de la educación difícilmente podríamos proceder de diferente manera, dado que antes, durante y después del dato cuantitativo se encuentra una realidad educativa de hechos y personas que es el objetivo último de la investigación. Y, es que, la investigación cualitativa, entendida como un análisis e interpretación minucioso, está en todo el proceso de investigación educativa.

MODELO

El modelo evaluativo común a las investigaciones que comentamos, podríamos identificarlo con el modelo CIPP de Stufflebeam y Shinkfield (1987), orientado a la valoración de los programas en su concepción, realización y resultados, para informar a quienes toman decisiones sobre cómo deben orientar la mejora de los programas y su gestión. Analizan variables de contexto, entrada, proceso y resultados. R. Pérez Juste utiliza su modelo de evaluación de programas educativos (inicial, procesual y final), que guarda similitudes con el de Stufflebeam y Shinkfield y el de la figura de Stake (1967).

METODOLOGÍA DE LA EVALUACIÓN

Los diseños que predominan son los cuasi-experimentales. El muestreo que se utiliza es diverso: incidental con grupos extremos, grupos naturales, muestreo aleatorio estratificado.

Las técnicas de recogida de información más utilizadas son las de tipo encuesta. Cuestionarios, entrevistas, observación. Se realizan análisis de documentos y se captura la realidad mediante instrumentos de audio y vídeo.

Los criterios para valorar la calidad de la información han sido pruebas de fiabilidad y validez, y su adecuación a las teorías.

Las referencias principales son la adecuación a los modelos, cuando los hay.

Las técnicas de registro y análisis de datos han sido, desde cuestionarios, escalas, pruebas académicas, observaciones mediante pautas e instrumentos técnicos, etc. ***Los análisis de los datos*** van desde los descriptivos, a los inferenciales. Predomina el análisis de la varianza y covarianza, contrastes mediante la prueba χ^2, análisis factoriales y discriminantes, entre otros.

El tipo de decisiones se orienta a recomendar las mejoras necesarias derivadas de las carencias y limitaciones que han sido detectadas en el proceso y resultado de las evaluaciones, dirigidas a aquellos que pueden realizarlas y a quienes tienen derecho a conocer los resultados de la evaluación y las propuestas de mejora como consecuencia de la misma.

Con mayor detalle, podemos ver la metodología de las siete investigaciones evaluativas que comentamos.

1) *Diseño.* Etnográfico. Metodología predominante cualitativa, de estudio de caso. Selección de cuatro Centros Asociados de la UNED, los dos de superior e infe-

rior rendimiento en el curso de acceso a la UNED para mayores de 25 años (CAD). Acceso al campo para analizar en profundidad las variables de estudio: el propio centro y su contexto, sus recursos, su organización, su profesorado, sus alumnos y los procesos instruccionales.
Técnicas de recogida de información. De diversas fuentes, técnicas e informantes. Análisis de documentos, entrevistas individuales y grupales, pruebas de conocimientos, observación directa, cuestionarios de opinión.
Criterios utilizados para valorar la información. Fiabilidad y validación de los instrumentos. Coherencia con las teorías e investigaciones previas.
Referencias. Teorías e investigaciones previas.
Técnicas de registro y análisis de la información. Análisis de las entrevistas, escalas de observación, cuestionarios. Análisis descriptivos, prueba de χ^2 entre los centros, análisis de la varianza de las pruebas de conocimientos; análisis cualitativos de los documentos, entrevistas y observación prolongada en los centros de sus directores, profesores, alumnos y personal administrativo.
Tipos de decisiones. Recomendaciones a la dirección de la UNED, —sede central y centros asociados— para una mayor incidencia en objetivos pedagógicos, mayor implicación de la dirección en la organización orientada al rendimiento y satisfacción de alumnos y profesores-tutores, mejora estructural de los espacios, y mayor implicación de la Junta de Gobierno de la UNED en la reducción de los desequilibrios presupuestarios entre los centros asociados de la UNED.

2) ***Diseño.*** Cuasi-experimental, jerárquico de seis grupos intratratamiento, transversal y longitudinal, a lo largo de dos cursos académicos (1989/1991); se realizado en análisis de variables estructurales y procesuales.
Técnicas de recogida de información. Observación sistemática, escalas, cuestionarios. Entrevistas, registro de incidentes. Reuniones con padres, profesores y otros implicados. Consulta de documentación.
Criterios utilizados para valorar la información. Accesibilidad, suficiencia y efectividad para medir las variables propuestas.
Referencias. Modelo teórico propuesto en el Plan Experimental para la Reforma de la Enseñanza Infantil (Marco teórico y Diseño Curricular Base).
Técnicas de registro y análisis de la información. Análisis de escalas de observación, cuestionarios y entrevistas. Análisis descriptivos, correlaciones, de covarianza y análisis de series temporales.
Tipos de decisiones. Formulación de propuestas organizativo-didácticas orientadas a mejorar el logro de los objetivos propuestos en el Plan Experimental.

3) ***Diseño*** no experimental: evaluación «ex-post-facto» del programa en cuanto tal, evaluación continua del proceso y de los resultados.
Técnicas de recogida de información. Diversas, en coherencia con el principio de complementariedad para la evaluación del programa: cuestionarios (expertos, alumnos y profesores). Evaluación del proceso: entrevistas en profundidad,

entrevistas telefónicas, cuestionario (alumnos y profesores). Evaluación de la eficacia: calificaciones finales (superación o no del curso).
Criterios utilizados para valorar la información. Evaluación del programa: adecuación a las demandas; utilidad; calidad de los objetivos y de los contenidos; calidad didáctica de los materiales. Evaluación del proceso: suficiencia y satisfactoriedad de los apoyos; dificultad y factibilidad de las actividades, satisfactoriedad por la información recibida y puntualidad de la misma. Eficacia en el logro de los objetivos.
Referencias. Sin referencia, al no haber datos comparables en la UNED ni haber planteado la Dirección unos niveles a alcanzar.
Técnicas de registro y análisis de la información. Análisis descriptivos. Comparación de las valoraciones de los diferentes informantes. Coherencia entre valoraciones de expertos, apreciaciones significativas (análisis textual).
Tipos de decisiones. Al no ser los autores de la evaluación los responsables del Programa, no cupo otra actuación que la elaboración del informe con propuestas de mejora.

4) ***Diseño.*** Cuasi-experimental de cuatro grupos independientes, con medidas pretest-postest. Acercamiento cualitativo a las situaciones iniciales y contextuales, y seguimiento del proceso.
Técnicas de recogida de información. Pruebas de rendimiento, cuestionarios de actitud, entrevista individual y grupal, antes y después de la experiencia.
Criterios utilizados para valorar la información. Adecuación del programa y de los instrumentos de recogida de información a la teoría y objetivos de la investigación, valorado por expertos. Durante el proceso: accesibilidad a la información mediante la herramienta del videotexto interactivo; seguimiento telefónico con los participantes en la experiencia: alumnos y profesores-tutores. Final: suficiencia y satisfactoriedad de los apoyos y de la información, y de los resultados académicos y motivacionales.
Referencias. Objetivos previamente fijados de rendimiento, motivacionales y actitudinales.
Técnicas de registro y análisis de la información. Cuestionarios de actitud y de rendimiento. Análisis descriptivos, prueba χ^2, análisis de la varianza en las pruebas de conocimientos; análisis cualitativos de entrevistas tras la experiencia.
Tipos de decisiones. Recomendaciones para introducir mejoras en la dirección del aprendizaje y evaluación formativa a través de los medios telemáticos en educación a distancia.

5) ***Diseño.*** Cuasi-experimental, pretest-postest, con grupos de clase naturales sin manipular: grupos experimental y control para alumnos de altas capacidades; grupos experimental y control para alumnos del promedio.
Técnicas de recogida de información. Protocolo para profesores por el que se seleccionan cinco alumnos de acuerdo con las características referenciadas. Pruebas: WPPSI, WISC, TTCT y Autoconcepto. Prueba de Pretest. Programa

piloto DASE y experimental DASE. Prueba de Postest. Recogida del rendimiento académico durante dos cursos académicos.
Criterios utilizados para valorar la información. Viabilidad, evaluabilidad, eficacia, eficiencia.
Referencias. Progreso sobre el propio programa.
Técnicas de registro y análisis de la información. Análisis de varianza, Análisis factorial para las actividades del programa DASE, índices de fiabilidad, validez y regresión. Prueba de χ^2.
Tipos de decisiones. El desafío intelectual del entorno de los alumnos de altas capacidades les induce a reducir esfuerzos, logrando buenos resultados con escasa inversión de tiempo y capacidad. El programa debe ser más específico y posiblemente incluir algunas sesiones de atención fuera del aula ordinaria y en el propio entorno familiar.

6) ***Diseño***. Cuasi-experimental. Primero: evaluación de las necesidades de orientación. Se estudian tres grupos diferentes pertenecientes a las once universidades de la Comunidad de Madrid: a) 2489 universitarios —muestreo aleatorio estratificado proporcional según tipos de carrera y universidades—, b) 44 orientadores y c) 11 responsables de los servicios de orientación. Segundo: análisis de los servicios de orientación de las instituciones universitarias; enfoque mixto (cualitativo y cuantitativo), valoraciones globales y estudio comparado de casos.
Técnicas de recogida de información. Cuestionarios para universitarios para profesionales. Entrevistas a los responsables de los equipos de orientación.
Criterios utilizados para valorar la información. Validación del cuestionario: La fiabilidad *test-retest*. Validez de constructo (análisis factorial para grupos de ítems que conforman una variable) y de contenido (revisión crítica de jueces).
Referencias. Las necesidades detectadas entre los universitarios junto a variables situacionales del alumnado y valoraciones aportadas por los agentes implicados. Los elementos comunes y divergentes apreciados en la comparación de los diversos servicios/programas.
Técnicas de registro y análisis de la información. En los análisis de las necesidades de orientación, estadísticos descriptivos y análisis de varianza, prueba χ^2 y análisis discriminantes. En el análisis de los servicios de orientación, estadísticos descriptivos y análisis de contenido de las respuestas de los universitarios, profesionales y directores a las entrevistas y la observación.
Tipos de decisiones. Destacando el importante desarrollo de los servicios de orientación universitaria, se detectan limitaciones que aconsejan la ampliación y mejora de los modelos de orientación, de sus objetivos (la actividad se centra en la función informativa y en el aspecto laboral/profesional, estando poco desarrolladas las funciones orientadora y formativa, académica y personal), y déficit en recursos materiales y personal especializado.

7) ***Diseño***. Cuasi-experimental, jerárquico, transversal de treinta y seis grupos intratratamiento. La complejidad de los fenómenos educativos a estudiar y la

falta de teorías potentes que proporcionen un marco previo consistente, recomiendan la utilización de un enfoque multiparadigmático en el que metodologías diversas se combinan.

Técnicas de recogida de información. Observación externa no participante, selectiva, descriptiva. Aplicación de escalas, cuestionarios y entrevistas.
Registro de incidentes. Análisis de grabaciones en audio y vídeo. Consulta de documentación. Reuniones con padres, profesores y otros implicados en la actividad escolar.

Criterios utilizados para valorar la información. Accesibilidad, suficiencia y efectividad para medir las variables propuestas.

Referencias. Marco teórico-conceptual derivado de la revisión y formulación del enfoque intercultural en educación: objetivos y procedimientos.

Técnicas de registro y análisis de la información. Aplicación y análisis de hojas de registro, escalas de observación, cuestionarios y entrevistas realizados. Análisis descriptivos, correlaciones, de series temporales. Representación gráfica mediante histogramas, diagramas de flujo y mapas conceptuales de la información contextual y procesual acumulada. Establecimiento del grado de convergencia entre datos obtenidos por diversas vías y fuentes.

Tipos de decisiones. Formulación de teorías más potentes desde las que orientar acciones educativas en contextos escolares multiculturales. Estrategias para ajustar las prácticas escolares al modelo de educación intercultural propuesto.

INFORME

1) Tesis doctoral. Informe al rector de la UNED. Informe abreviado a los directores de los centros asociados. Difusión en congresos. Publicación: Martínez Mediano, C. (1993): *Factores de eficacia de los Centros Asociados de la UNED*. Madrid: UNED. ICI AIESAD. Martínez Mediano, C. (1994): Factores de eficacia de los Centros Asociados de la UNED. *Revista de la AIESAD*. Martínez Mediano, C. (1996): Investigación evaluativa sobre los factores de eficacia de los Centros Asociados de la UNED. En L. García Aretio: *Investigar para mejorar la calidad de la UNED. Trabajos de investigación sobre la UNED*. Madrid: IUED.UNED.
2) Tesis doctoral. Publicación: Aguado, T. (1993): *Análisis y evaluación de modelos y programas de educación infantil en el marco de la Reforma de la Enseñanza*. Cuadernos de la UNED, 124, Madrid.
3) Informe de investigación, de 168 páginas, entregado al Rector de la UNED a la Dirección General de Tráfico y a la Dirección del Programa.
4) Informes y difusión en encuentros científicos de la Institución. Publicado en actas de congresos y revistas. Varios (1997): Evaluación formativa de los alumnos de Pedagogía Experimental (adaptación) a través del videotexto interactivo. En C. Martínez Mediano, (Coord.). *Encuentros en la Facultad de Educación sobre Evaluación*. Madrid: UNED. Pérez Juste, R. y Sebastián, A. (1992): «Using videotex for formative evaluation in UNED (Spain)». En **Conference Abstracs. Distante Education for the Twenty first century.** International Council for Distante

Education. Sukhothai Thammathirat Open University. Pérez Juste. R. (1993): **Papel nuclear de la evaluación en la calidad y eficacia de los sistemas educativos a distancia,** Ponencia de la Sección III del Congreso Internacional *La Educación superior a Distancia: Perspectivas de futuro.* Valdepeñas. Pérez Juste, R (en colaboración) (1993): «Incidencia de la dirección del aprendizaje por ordenador y de la evaluación formativa en los resultados del aprendizaje, en la satisfacción en el estudio y en la actitud hacia la institución universitaria», Comunicación al Congreso Internacional *La Educación superior a Distancia: Perspectivas de futuro.* Valdepenas.

5) Tesis doctoral. Publicación: Álvarez González, B. (1998): *Un Programa de Enriquecimiento para alumnos de cinco a siete años.* Madrid: Premios CIDE de investigación educativa. Álvarez González, B. (2000): Alumnos de altas capacidades. Identificación e intervención educativas. Madrid: Bruño.
6) Tesis doctoral. Publicación: Sánchez García, M.F. (1999): *Necesidades y servicios de orientción universitaria en la Comunidad de Madrid.* Madrid: UNED. Destinado a los responsables y profesionales de los servicios de orientación y a las autoridades académicas, a fin de que en las políticas educativas universitarias se tenga en cuenta la necesidad de incorporar programas y servicios de orientación.
7) Como informe de investigación ante el CIDE, ya que se realizó gracias a una de sus convocatorias de ayudas a la investigación. Publicación: Aguado, T. y cols. (1999): *Diversidad cultural e igualdad escolar. Un modelo para el diagnóstico y desarrollo de actuaciones educativas en contextos multiculturales.* CIDE/MEC, Col. Investigación, 141, Madrid.

695

A.I.D.I.P.E.
BOLETÍN ASOCIADO

Para asociarse llene las dos partes de este boletín y devuélvalo a:

A.I.D.I.P.E.
Dpto. «Métodos de Investigación y Diagnóstico en Educación»
Facultad de Pedagogía **Paseo del Valle de Hebrón, 171, 2ª planta**
08035 - BARCELONA (Spain)
Cuota de suscripción anual 6.000 ptas.

DATOS PERSONALES

Nombre y Apellidos .....
D.N.I. o N.I.F. .....
Dirección .....
Población ..... C.P.....
Provincia ..... E-mail:..... Teléfono ( ).....
Deseo asociarme desde el día ..... de..... de 20.....
DPTO. TRABAJO..... **CENTRO TRABAJO**.....
Situación profesional ..... Dist. Universitario.....

DATOS BANCARIOS

Titular de la cuenta .....
Banco/Caja .....
Domicilio Agencia .....
Población ..... C.P.....

(Firma)

CODIGO CUENTA CLIENTE			
Entidad	**Oficina**	**D.C.**	**Núm. de Cuenta**

A.I.D.I.P.E.
BOLETÍN DOMICILIACIÓN BANCARIA

Señores,
Les agradeceré que con cargo a mi cuenta/libreta atiendan los recibos que les presentará la **Asociación Interuniversitaria de Investigación Pedagógica** como pago de mi cuota de asociado.

Titular de la cuenta .....
Banco/Caja .....
Domicilio Agencia .....
Población ..... C.P.....

(Fecha y Firma)

CODIGO CUENTA CLIENTE			
Entidad	**Oficina**	**D.C.**	**Núm. de Cuenta**

REVISTA INVESTIGACIÓN EDUCATIVA

BOLETÍN SUSCRIPCIÓN

Para subscribirse llene este boletín y devuélvalo a:
REVISTA DE INVESTIGACIÓN EDUCATIVA
Dpto. «Métodos de Investigación y Diagnóstico en Educación»
Facultad de Pedagogía Paseo del Valle de Hebrón, 171, 2ª planta
08035 - BARCELONA (Spain)

Nombre .....

D.N.I. o N.I.F. .....

Dirección .....

Población ..... C.P. .....

País ..... E-mail:..... Teléfono.....

Coste de la inscripción:

- ☐ Individual: 6.000 + I.V.A. ptas.
- ☐ Institucional: 10.000 + I.V.A. ptas.
- ☐ Números sueltos: 2.500 ptas.
- ☐ Indicar n.º deseado: .....
- ☐ Números extras y monográficos: 3.000 ptas.
- ☐ Indicar n.º deseado: .....

SUSCRIPCIÓN INTERNACIONAL

INDIVIDUAL

- ☐ Europa: 6.000+2.000 ptas. gastos de envío
- ☐ América: 6.000+3.000 ptas. gastos de envío

(Fecha y Firma)

INSTITUCIONAL

- ☐ Europa: 10.000+2.000 ptas. gastos de envío
- ☐ América: 10.000+3.000 ptas. gastos de envío

BOLETÍN DE DOMICILIACIÓN BANCARIA

Señores,

Les agradeceré que con cargo a mi cuenta/libreta atiendan los recibos que les presentará la *Revista de Investigación Educativa*, como pago de mi subscripción a la misma.

Titular de la cuenta .....

Banco/Caja .....

N.º de cuenta ..... N.º de libreta .....

Agencia .....

Población .....

(Fecha y Firma)

AIDIPE

Asociación Interuniversitaria de Investigación Pedagógica, creada en 1987

FINES DE LA ASOCIACIÓN

a) Fomentar, estimular e impulsar la investigación educativa de carácter empírico en todas sus vertientes y ámbitos de aplicación.
b) Difundir los resultados de esta investigación.
c) Agrupar a todos los que en España cultivan esta área de conocimiento, como docentes o investigadores.
d) Colaborar con las autoridades educativas y de política científica en la elaboración de instrumentos de medida, evaluación y de diseños de investigación tendentes a la mejora del sistema educativo y de desarrollo de los mismos.
e) Favorecer la formación y el perfeccionamiento especializado del personal investigador y de todos aquellos profesionales que intervienen en la investigación educativa.
f) Velar por la calidad científica y profesional de las actividades relacionadas con la investigación educativa.
g) Velar por el respeto y cumplimiento de las normas deontológicas de la investigación científica referida al campo psicopedagógico.

DERECHOS DE LOS SOCIOS

Los socios numerarios y fundadores tendrán los siguientes derechos:

a) Tomar parte en cuantas actividades organice o patrocine la Asociación en cumplimiento de sus fines.
b) Disfrutar de todas las ventajas y beneficios que la Asociación pueda obtener.
c) Participar en las Asambleas con voz y voto.
d) Ser electores y elegibles para los cargos directivos.
e) Recibir información sobre los acuerdos adoptados por los órganos directivos, sobre las actividades técnicas y científicas que la Asociación desarrolle.
f) Hacer sugerencias a los miembros de la Junta Directiva de la Asociación.
g) Impugnar los acuerdos y actuaciones de la Asociación que sean contrarios a los Estatutos, dentro del plazo de cuarenta días y en la forma prevista por las leyes.

Cada socio tiene derecho a recibir la Revista de Investigación Educativa de forma gratuita.

SEMINARIOS

AIDIPE organiza cada 2 años, un Seminario Nacional con un tema central en torno al cual giran las ponencias, comunicaciones, mesas redondas. Los primeros Congresos han sido los de Barcelona (1981), Sitges (1983), Gijón (1985), Santiago de Compostela (1988), Murcia (1990).

PUBLICACIONES

AIDIPE publica las Actas de sus Seminarios, así como la Revista de Investigación Educativa (RIE) con carácter semestral.

Para más información, ver hoja de inscripción adjunta.

NORMAS PARA LAS COLABORACIONES

El objetivo de la Revista de Investigación Educativa es promover el intercambio de información acerca de investigaciones empíricas de carácter educativo. Todo profesional que desee colaborar en la revista deberá atenerse a las siguientes indicaciones:

1. Los trabajos deberán ser originales y versar sobre investigación educativa.
2. El autor o autora deberá enviar un original y tres copias mecanografiadas a la redacción de la Revista.
3. El autor o autora deberá especificar debajo del nombre del artículo su dirección profesional y su E.mail. Si son varios autores/as, se adjuntará el E.mail del primer autor como mínimo.
4. La extensión máxima de los trabajos no deberá exceder las 25 páginas en DIN-A4, a doble espacio y numeradas. Se acompañará un abstract de 100 a 150 palabras en inglés y español.
5. Con la finalidad de simplificar el proceso de confección de la revista y, sobre todo, de reducir al máximo las erratas y costos, se ruega a los autores/as enviar el trabajo, además de las copias en papel, en disquette compatible. Se aceptan los siguientes programas de tratamiento de textos: Word Star, Word Perfect, MSWord y otros.
6. Los trabajos recibidos serán sometidos a informe del Consejo Asesor de la Revista (quien decidirá sobre su publicación o no).
7. Los trabajos recibidos para su publicación en la revista estarán sujetos a un doble proceso de filtro para garantizar la objetividad del juicio. En primer lugar, existirá un proceso de separación de las identificaciones de los autores/as de los trabajos candidatos a ser publicados junto con la asignación de códigos identificativos. Además, existirá un proceso completamente independiente de elección de jueces entre el Consejo Asesor. Ambos procesos serán totalmente independientes y transparentes entre sí, de forma que al asignar un conjunto de jueces del Consejo Asesor se desconoce cualquier información sobre el artículo, salvo el conjunto de descriptores sobre su contenido.
8. Se notificará a sus autores/as los trabajos aceptados para su publicación.
9. Para la redacción de los trabajos se recomienda que sigan las normas de la A.P.A. (American Psychology Association). (Ver adaptación publicada en el n.º 19, 1.er semestre 1992 de R.I.E.).
10. Las fichas resumen se cumplimentarán de acuerdo con el modelo propuesto e impreso por la Revista, a cuya sede podrá solicitar los ejemplares.
11. Los números se cierran los días 1 de mayo y 1 de noviembre de cada año.
12. La R.I.E. es de carácter semestral.
13. Los autores/as deben incluir en sus artículos referencias a otros trabajos publicados en RIE que traten sobre el mismo tema.

REDACCIÓN: REVISTA DE INVESTIGACIÓN EDUCATIVA
Dpto. de Métodos de Investigación y Diagnóstico en Educación
Facultad de Pedagogía
08035 BARCELONA (Spain)

Sumario

Presentación 251
Ramón Pérez Juste

PRIMERA PARTE

La evaluación de programas educativos: conceptos básicos, planteamientos generales y problemática 261
Ramón Pérez Juste

La evaluación de programas sociales. Fundamentos y enfoques teóricos 289
Mario de Miguel Díaz

El diseño y los diseños en la evaluación de programas 319
Fco. Javier Tejedor Tejedor

La validez en la evaluación de programas 341
J.M. Jornet Meliá, J.M. Suárez y A. Pérez Carbonell

El análisis de datos en la evaluación de programas educativos 357
Luis Lizasoain Hernández y Luis Joaristi Olariaga

La función optimizante de la evaluación de programas evaluativos 381
Arturo de la Orden Hoz

SEGUNDA PARTE

El muestreo y su problemática en las evaluaciones de programas institucionales 393
José Luis Gaviria Soto

La evaluación y mejora de la enseñanza en la Universidad: otra perspectiva 405
Tomás Escudero Escorza

La evaluación de la actividad docente en la Universidad: entre el sueño y la realidad 417
Gregorio Rodríguez Gómez

La evaluación de la enseñanza por los alumnos en el Plan Nacional de Evaluación de la Calidad de las Universidades. Construcción de un instrumento de evaluación 433
Juan Ruiz Carrascosa

Evaluaciones de programas de formación universitaria en el ámbito europeo y americano. Coherencia con las demandas sociales 447
Honorio Salmerón Pérez

Nuevas tendencias en la evaluación de programas de Educación Multicultural 463
Margarita Bartolomé Pina y Flor Cabrera Rodríguez

La evaluación de programas informatizados 481
Víctor Álvarez Rojo, Eduardo García Jiménez, Javier Gil Flores, Soledad Romero Rodríguez y Javier Rodríguez Santero

El programa de orientación «Tu futuro profesional» y su evaluación 493
Elvira Repetto y J. Antonio Gil

El orientador como investigador de la acción 509
Rafael Sanz Oro

La evaluación de la formación ocupacional y continua 521
J.M. Jornet, J.M. Suárez y Mª J. Perales

Evaluación de programas para la educación de sobredotados 539
José Cajide Val

Evaluación de programas para alumnos superdotados 553
Carmen Jiménez Fernández

Evaluación de programas para alumnos de alta capacidad: Algunos problemas metodológicos 565
Javier Tourón

Evaluación de programas de educación emocional 587
Manuel Álvarez, Rafael Bisquerra, Eva Fita, Francesc Martínez y Nuria Pérez

Evaluación de programas de atención a los niños con discapacidades 601
María Luisa Dueñas

TERCERA PARTE 611

Universidades de: Alicante, Barcelona, Cádiz, Complutense, Granada, Jaén, Málaga, Murcia, Navarra, País Vasco/EHU, Salamanca, Santiago de Compostela, Sevilla, Nacional de Educación a Distancia

Web: Rie www.um.es/~dipmide
Aidipe www.uv.es/aidipe

700

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").